UNIVERSITAT ROVIRA I VIRGILI

EMPOWERING COGNITIVE STIMULATION THERAPY WITH SOCIALLY ASSISTIVE ROBOTICS AND EMOTION RECOGNITION

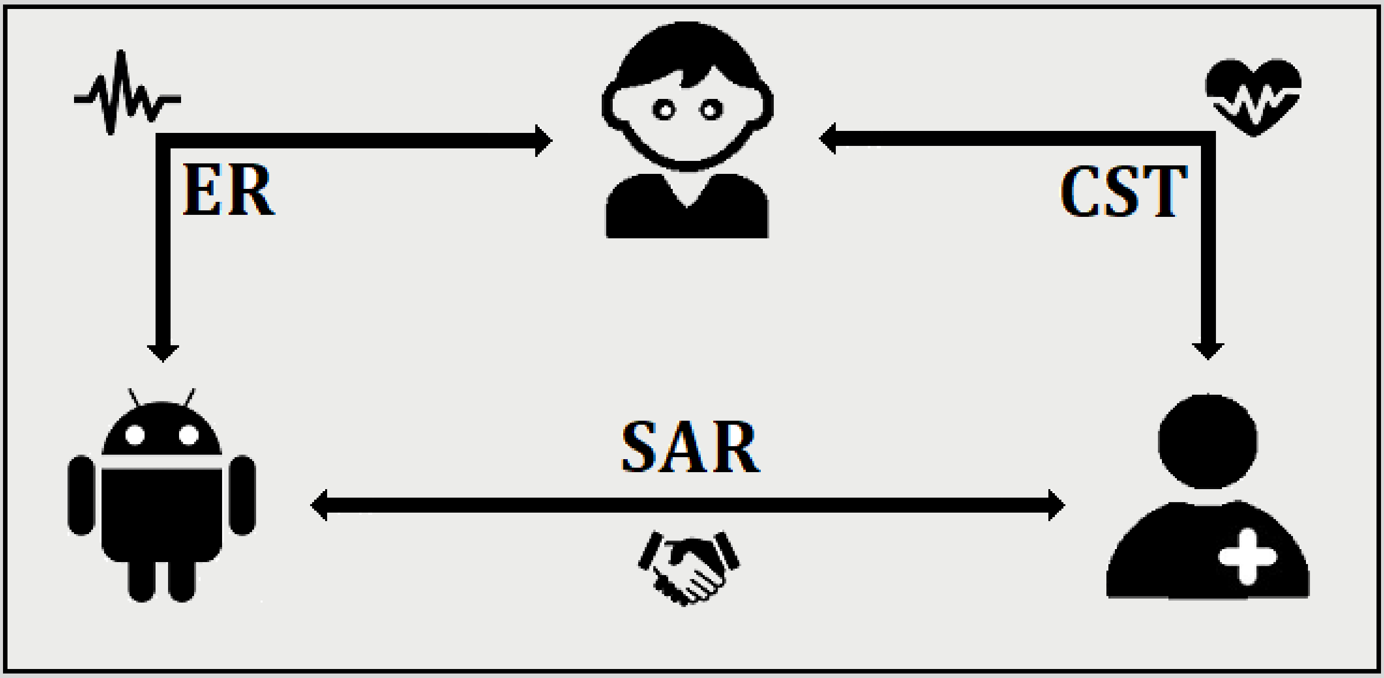
Jainendra Shukla

UNIVERSITAT
ROVIRA i VIRGILI

Empowering Cognitive Stimulation Therapy with Socially Assistive Robotics and Emotion Recognition

---

JAINENDRA SHUKLA



DOCTORAL THESIS
2018

Amb el suport de la Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya.

Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca

DOCTORATS INDUSTRIALS

AGAUR
Agència de Gestió d'Ajuts Universitaris i de Recerca

Instituto de Robótica PARA LA DEPENDENCIA
FUNDACIÓN PRIVADA

Empowering Cognitive Stimulation Therapy with Socially Assistive Robotics and Emotion Recognition

UNIVERSITAT ROVIRA I VIRGILI

Jainendra Shukla

Departament d'Enginyeria Informàtica
i Matemàtiques

University Rovira i Virgili (URV)

DOCTORAL THESIS

Supervised by Prof. Domènec Savi Puig Valls

Tarragona May 2018

UNIVERSITAT
ROVIRA I VIRGILI
**Departament d'Enginyeria Informàtica
i Matemàtiques**

Av. Paisos Catalans, 27
43007 Tarragona
Tel. +34 977 55 95 95
Fax. +34 977 55 95 97

I STATE that the present study, entitled "Empowering Cognitive Stimulation Therapy (CST) with Socially Assistive Robotics (SAR) and Emotion Recognition", presented by Jainendra Shukla, for the award of the degree of Doctor, has been carried out under my supervision at the Departament d'Enginyeria Informàtica i Matemàtiques of this university, and that it fulfills all the requirements to be eligible for Industrial Doctorate Distinction and International Doctor Distinction Award.

Tarragona, 21st March 2018

Doctoral Thesis Supervisor

Dr. Domènec Savi Puig Valls

॥ श्रीसीतारामार्पणमस्तु ॥

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. I have made full acknowledgements of the work and ideas of any other people who are cited in the thesis, or who have contributed to it.

Jainendra Shukla
May 2018

Acknowledgements

॥ नमो राघवाय॥

Completion of this thesis would not have been possible without the invaluable help and support of a number of people.

Thanks to the Industrial Doctorate program (Ref. ID.: 2014-DI-022) of AGAUR, Govt. of Catalonia for funding this research. I am glad to pursue my doctoral studies under Industrial Doctorate program and am thankful to the creators of this program. I would like to extend my personal gratitude to Prof. Domènec Puig, my thesis supervisor at URV and Mr. Antoni Reverter for believing in me. I also would like to thanks Er. Joan Oliver, my colleague and research manager at IRD for his sustained interest, support and association. Thank you Miguel and Julián for your great collaborations and related discussions. I also would like to thank Prof. G.C. Nandi, my graduate supervisor for providing motivation and encouragement to pursue the doctoral studies. I deeply appreciate all the members of the Fundació Ave Maria, Sitges and Instituto de Robótica para la Dependencia, Sitges for their general support and friendship. I am also thankful to several anonymous reviewers for their valuable comments and suggestions which helped me to improve the quality of my publications and the thesis.

Thanks to my father and my mother for their loving and endless support. A special thanks to my Didi and my younger sister for being there when I needed them most. I also would like to thanks my lovely fiancée Deepika for her support which started as a friend and will now continue as my better half. I also would like to thanks my Vaishnava friends for their beautiful personal and spiritual association over the several years.

Finally, thank you Shri SadGurudeva Bhagwan and Lord Shri SitaRam for your causeless mercy and absolute compassion.

॥ श्रीराघवः शंतनोतु॥

Abstract

Socially Assistive Robotics (SAR) has already been widely used in mental health service and research, primarily among children with Autism Spectrum Disorder (ASD) and among older adults with dementia. Motivated by the benefits offered by SAR in mental health service and research, we envision that SAR can also benefit cognitive rehabilitation of individuals struggling with a wide range of mental health concerns, including adults with intellectual developmental disorders (IDD), people with neurodegenerative disorders such as Alzheimer's disease etc. Cognitive rehabilitation involves guided practice on a range of standard tasks related to one or more cognitive domains. Such gain in cognition will increase autonomy among these individuals which in turn can improve quality of life and hence well-being of these individuals. Motivated by lack of adequate resources for providing support to individuals with mental health concerns, the benefits offered by cognitive rehabilitation and SAR in mental health service and research, we envision that SAR empowered cognitive rehabilitation can positively affect the well-being of a wide variety of users.

In this thesis, we investigated the benefits of robot assisted cognitive rehabilitation for individuals with IDD. In the first part, we evaluated the fitness of robot assisted interventions for cognitive rehabilitation of individuals with IDD. We conducted interviews with seven expert psychologists and professional caregivers working with individuals with IDD. These interviews helped us to identify key aspects of a beneficial robot-assisted mental health intervention. To assess the impact of such robot-assisted mental health interventions on the users, we conducted a case study of robot interactions among six individuals with IDD using NAO robot in different categories of interaction. We also compared the response of robotic interactions with non-robotic visual stimulations caused by a tactile gaming console. The results reported positive effects of robot interventions on the users and that the stimulations caused by tactile gaming console can significantly serve as complementary tool for therapeutic benefit of patients. We further evaluated the impact of robot-assisted mental health interventions on caregivers in multi-center trials. The results of the research confirmed a significant reduction in caregivers burden during execution of cognitive stimulation interventions for individuals with IDD and raised a concern about the need of a specific training of the caregivers to take maximum advantage of SAR in health care.

The second part concerns itself with empowering the social robots with automated and online emotion recognition ability for emotional adaptation in a bid to improve rehabilitation. We conducted a series of cognitive stimulation sessions among individuals with IDD in a nearly real world settings to obtain a first ever annotated multimodal dataset (MuDERI) of individuals with IDD. MuDERI is an annotated multimodal dataset of audiovisual recordings, RGB-D videos and Electro-dermal activity (EDA), Electroencephalogram (EEG) physiological signals of 12 participants in actual settings, which were recorded as participants were elicited using personalized cognitive stimulation sessions. The dataset is publicly available. We further proposed an efficient wavelet-based method for artifacts attenuation of EDA signal during the online collection and analysis, while minimizing distortions, using a stationary wavelet transform (SWT). The proposed method was tested on EDA recordings from publicly available driver dataset collected during real-world driving, and containing a high number of motion artifacts, and the results were compared to those of three state-of-the-art methods for EDA signal filtering. In addition, the proposed method was tested for the online filtering of EDA signals collected while twelve volunteers conducted tasks designed to elicit various stress states. The results evidenced that the prediction of arousal states can be significantly improved after motion artifacts removal, and that the proposed method outperforms existing approaches and it has a lower computational cost. Taken together, these results evidence the effectiveness of the proposed method for online EDA filtering in real world scenarios. Further, we reviewed feature extraction methods for emotion recognition from EDA based on 25 studies. We compared these features for feature selection using machine learning techniques on a publicly available AMIGOS dataset. We present the results of the performance of three feature selection methods and usage of selected feature types across time, frequency, time-frequency domains. We did not find any statistical evidence that any of the three employed feature selection methods outperform the others. However, the subject-dependent classification results were significantly higher than the subject-independent classification for both the arousal and the valence recognition. Mel-frequency cepstral coefficients (MFCC) and related statistical features were explored for the first time for the emotion recognition from EDA signals and they outperformed all other feature types, including the most commonly used Skin Conductance Response (SCR) related features. We also compared our results with methods employed by researchers of AMIGOS dataset for classification of emotional states and they show that the EDA features explored in this study provided better performance, validating the findings of our study.

Keywords: Socially Assistive Robotics, Intellectual Disability, Robot Interaction, Rehabilitation, SAR,IDD.

Table of contents

List of figures

List of tables

Nomenclature

Greek Symbols

δ	Proportion of Motion Artifacts in Original Signal
γ	Mixture Parameter for Gaussian Mixture Model
μ	Location Parameter of Laplace Distribution
$\hat{\mu}$	Maximum Likelihood Estimator of Location Parameter
Φ	Cumulative Distribution Function of the Laplace Distribution
σ	Level of Noise

Subscripts

j	subscript index

Other Symbols

b	Scale Parameter of Laplace Distribution
$\hat{b}$	Maximum Likelihood Estimator of Scale Parameter

Acronyms / Abbreviations

ASD	Autism Spectrum Disorder
CA	Cepstrum Analysis
CMIM	Conditional Mutual Information Maximization
CST	Cognitive stimulation therapy
DAQ	Data Acquisition

dbN	Daubechies wavelets
$DISR$	Double Input Symmetrical Relevance
$DTFT$	Discrete-Time Fourier Transform
DWT	Discrete Wavelet Transform
EDA	Electro-dermal activity
EEG	Electroencephalogram
EMG	Electromyography
FAM	Fundació Ave Maria
FFT	Fast Fourier Transform
$freq$	Frequency
DWT	Gaussian Mixture Model
HF	High Frequency
HMI	Human Machine Interaction
HOC	Higher Order Crossing
HRI	Human Robot Interaction
Hz	Hertz
IAB	Institutional Advisory Board
ICD	International Classification of Diseases
ID	Intellectual Disability
IDD	Intellectual Developmental Disorders
$IDTFT$	Inverse Discrete-Time Fourier Transform
J	Level of Wavelet Decomposition
JMI	Joint Mutual Information
log	Logarithm

MAP Maximum a Posteriori

MFCC Mel-Frequency Cepstral Coefficients

$NASA-TLX$ NASA Task Load Index

PSD Power Spectra Density

RFID Radio-frequency identification

$RGB-D$ Red Green Blue-Depth

SAR Socially Assistive Robotics

SCL Skin Conductance Level

SCR Skin Conductance Response

SMNA Sudomotor Nerve Activity

STFT Short-Time Fourier Transform

SWT Stationary wavelet transform

WHO World Health Organization

Chapter 1

Introduction

1.1 Motivation

Socially Assistive Robotics (SAR) has already been widely used in mental health service and research (Rabbitt et al., 2015), primarily among children with Autism Spectrum Disorder (ASD) (Coeckelbergh et al., 2016) and among older adults with dementia (Moyle et al., 2013). This work is a holistic approach to extend the benefits of SAR for cognitive rehabilitation of Individuals with Intellectual Disability (ID).

As per the latest proposition made by the International Classification of Diseases (ICD) Working Group of the World Health Organization (WHO), mental retardation/intellectual disability terms have been replaced with the concept of *Intellectual Developmental Disorders (IDD)* and is defined as *"A group of developmental conditions characterized by significant impairment of cognitive functions, which are associated with limitations of learning, adaptive behavior and skills"* (Carulla et al., 2011). The diagnostic term *Intellectual Disability (ID)* is the equivalent term for the ICD-11 diagnosis of *Intellectual Developmental Disorders (IDD)*, but Intellectual Disability is still the commonly used term in research journals and by professionals in the field (Association, 2013). For above reason, the term Intellectual Disability (ID) will be used throughout this thesis. Main characteristics of individuals with ID are (Carulla et al., 2011):

- Difficulties with verbal comprehension, perceptual reasoning, learning (including academic and practical knowledge), working memory, and processing speed due to extremely delayed/limited intelligence.
- Limited social and practical skills, which hinder their ability to effectively function in everyday lives. These skills include communication, social interactions, taking care of oneself, managing money, and using transportation among others.

- ID originates before the age of 18 (Association, 2013) and is a lifespan condition (Carulla et al., 2011).

ID affects about 1% to 3% of the world population (Organization, 2001). According to a recent study, the number of individuals with ID has increased from 18.2 million in 1990 to 154.0 million in 2013 (Vos et al., 2015). It is estimated that global burdens caused by these disabilities will increase to 15%, by 2020 (Organization, 2001). Researchers have not found any promising treatments for individuals with ID (Brown et al., 2013); as a result, ID remains a lifespan condition and thus providing support to individuals with ID demands a huge physical, economical and emotional support by the worldwide community.

Notably, cognitive rehabilitation can benefit cognition of individuals struggling with a wide range of clinical concerns, including adults with intellectual developmental disorders (IDD) (Peñaloza-Salazar et al., 2015), people with neurodegenerative disorders such as Alzheimer's disease (Worthington, 2005). Such gain in cognition will increase autonomy among these individuals which in turn can improve quality of life and hence well-being of these individuals (Worthington, 2005). Likewise, SAR has a wide application domain that includes care for the elderly, care for individuals with physical recovery/rehabilitation and training needs, and care for individuals with cognitive and social disabilities (Tapus et al., 2007). While SAR has not offered any permanent cure to any of the above mental concerns, it has shown a significant potential in increasing the quality of life among individuals affected by these diseases (Kanamori et al., 2003; Kozima et al., 2007; Pennisi et al., 2015; Tapus et al., 2009a).

Consequently, motivated by the benefits offered by cognitive rehabilitation and by SAR in mental health service & research and lack of adequate resources, we envision that SAR empowered cognitive rehabilitation can positively affect the well-being of a wide variety of users.

The work presented in this thesis was developed under the frame of the REHABIBOTICS project. REHABIBOTICS project aims to increase the quality of life among individuals with ID using robotics. It is a collaborative research between Intelligent Robotics and Computer Vision Group (IRCV) at Rovira i Virgili University, Tarragona, Spain and Instituto de Robótica para la Dependencia, Sitges, Spain. Project REHABIBOTICS will be deployed at Ave Maria Foundation (FAM), which is a residential care facility for individuals with ID[1].

[1] Ave Maria Fundació, http://www.avemariafundacio.org/inici.html

1.2 Problem statements and challenges

The ultimate goal of research in empowering SAR for cognitive rehabilitation is to make the robots "as effective" as an *experienced human caregiver*. While the use of SAR can not replace caregivers at residential care facilities, it can help them to focus their time and resources on providing better monitoring and personalized attention of the individuals. To make SAR capable of "effectively" assisting a human caregiver, it is important to understand that in this interdisciplinary field of research, the challenges arise from both an engineering and cognitive rehabilitation point of view. Addressing the issues in this way only, will lead to a meaningful advancement in the field. Fundamental challenges which directed the research and development efforts of current thesis are summarized below.

Cognitive Rehabilitation Challenges

It is important to ensure that the developed SAR empowered cognitive rehabilitation system will meet the end-user needs and is usable. Accordingly all the stakeholders must be actively involved in the requirements gathering phase to ensure its user-friendliness and usability. The stakeholders mainly are the end-users of the solution i.e. individuals with ID, their caregivers and the residential care facilities. While user studies for the development of the SAR systems have been done previously for different populations such as among kids with ASD, among elderly people with Dementia etc. But their implications on development of SAR empowered cognitive rehabilitation system is limited. Particularly, the impact of such robot-assisted interventions on individuals with ID and their caregivers has not received sufficient attention.

Engineering Challenges

From an engineering point of view, it is important to understand what types of robot capabilities will be required to achieve an active participation from the involved users. Also, it is necessary to identify the robotic features that can improve the user's task performance in such interventions. Such capabilities and features will lead to "natural" human-robot interaction, which has been limited by the technological advancements in the field. In particular, the robots do not have the ability to understand individual social signals and cues and to respond appropriately.

The aims of this dissertation are twofold:

- Evaluate the fitness of robot assisted interventions for cognitive rehabilitation among individuals with ID, while assessing its impact on the users and their caregivers.

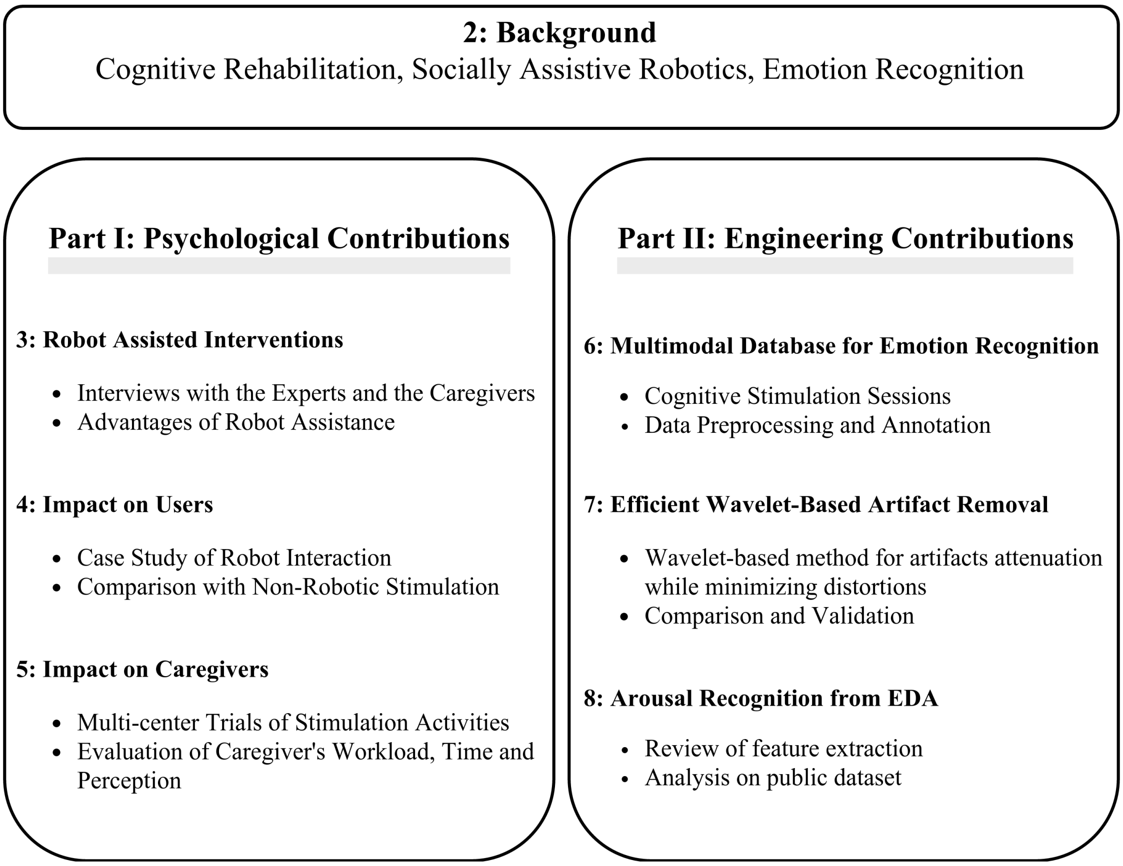


Fig. 1.1 Outline of the thesis.

- Develop automated and online emotion recognition ability for emotional adaptation using physiological signals.

1.3 Contributions and thesis outline

The thesis is structured as outlined in Fig. 1.1. To begin with, chapter 2 introduces fundamental concepts of cognitive rehabilitation, SAR and emotion recognition. Further, the main content is presented in chapters 3 through 8 and is divided into two parts. With respect to the problems identified in subsection 1.2, the contributions made within this thesis advance the state of the SAR empowered cognitive rehabilitation by addressing engineering challenges. Altogether, the main contributions of the thesis are presented in chapters 3 to 8 and can be divided into two parts:

Contributions Related to Cognitive Rehabilitation

- In the first part of the thesis we take a closer look at identifying the fitness of robot assisted interventions for cognitive rehabilitation of individuals with ID. To evaluate the fitness of robot assisted mental health interventions, we conducted interviews with expert psychologists and professional caregivers working among individuals with ID. These interviews helped us to identify different interaction categories, designing strategy for the above types of activities, different metrics for evaluation of such robot assisted interventions and we finally proposed an interactive and adaptive architecture of SAR assisted mental health interventions.

- We performed a case study of robot interactions among individuals with ID to assess the impact of robot assisted interventions on the users. This case study reported positive effects of such interventions on the users and helped us to verify the response of the users in front of the robot in different intervention scenarios. We also compared the response of robotic interactions with stimulations caused by a tactile gaming console, among individuals with ID. The results show that robot interactions are more effective but stimulations caused by tactile gaming consoles can significantly serve as complementary tool for therapeutic benefit of the patients.

- To assess the impact of SAR during cognitive stimulation therapy (CST) interventions on the caregivers, we conducted a cognitive training activity involving individuals with ID and their caregivers in multi-center trials. We conducted a multidimensional evaluation of caregivers workload; including subjective workload, time spent on users personalized interventions, and qualitative interviews with caregivers. The results of the research confirm a significant reduction in caregivers burden and raise a concern about the need of a specific training of the caregivers to take maximum advantage of SAR in health care.

Contributions Related to Engineering Challenges

- The second part concerns itself with achieving online emotion recognition among robots. We conducted a series of cognitive stimulation sessions among individuals with ID in a nearly real world settings to obtain a first ever annotated multi-modal database of individuals with ID. Two cognitive stimulation sessions were designed for each user, one aimed to elicit positive emotions (joy), and the other aimed to elicit negative emotions (sadness or anger) to create MuDERI[2] dataset. MuDERI

[2]https://institutorobotica.org/en/investigation/muderi-dataset/

is an annotated multimodal dataset of audiovisual recordings, RGB-D videos and Electro-dermal activity (EDA), Electroencephalogram (EEG) physiological signals of 12 participants in actual settings, which were recorded as they were elicited using personalized cognitive stimulation sessions. The dataset is publicly available.

- We propose an efficient wavelet-based method for artifacts attenuation while minimizing distortions, using a stationary wavelet transform (SWT). The proposed method was tested further on EDA recordings from publicly available driver dataset collected during real-world driving, and previously generated MuDERI dataset, and the results were compared to those of three state-of-the-art methods for EDA signal filtering. The results evidenced that the prediction of arousal states can be significantly improved after motion artifacts removal by the proposed method, and that it outperforms existing approaches and has a lower computational cost.
- We reviewed feature extraction methods for emotion recognition from EDA based on 15 studies. We compared these features for feature selection using machine learning techniques on a publicly available dataset.

1.4 Peer Reviewed Publications

Some of the results of this dissertation have appeared in the following publications:

Patent

1. Shukla, J., Oliver, J., and Puig, D. (2017b). A computer-implemented method for the measurement of human emotion of a subject and a method for filtering an eda signal. European Patent Application Number EP17382661, Submitted

Journals

2. Shukla, J., Barreda-Ángeles, M., Oliver, J., and Puig, D. (2018b). Efficient wavelet-based artifact removal for electrodermal activity in real-world applications. *Biomedical Signal Processing and Control*, 42C:45 – 52
3. Shukla, J., Barreda-Ángeles, M., Oliver, J., Nandi, G. C., and Puig, D. (2018a). Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Transactions on Affective Computing*, Submitted

4. Shukla, J., Cristiano, J., Oliver, J., and Puig, D. (2018c). Robot assisted interventions for individuals with intellectual disabilities: Impact on users and caregivers. *International Journal of Social Robotics*, Submitted

Book Chapters

5. Shukla, J., Barreda-Ángeles, M., Oliver, J., and Puig, D. (2016a). Muderi: Multimodal database for emotion recognition among intellectually disabled individuals. In *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016*, pages 264–273. Springer International Publishing, Cham

6. Shukla, J., Cristiano, J., Anguera, L., Vergés-Llahí, J., and Puig, D. (2016b). Robot 2015: Second iberian robotics conference: Advances in robotics, volume 2. chapter A Comparison of Robot Interaction with Tactile Gaming Console Stimulation in Clinical Applications, pages 435–445. Springer International Publishing, Cham

7. Shukla, J., Cristiano, J., Amela, D., Anguera, L., Vergés-Llahí, J., and Puig, D. (2015). Social robotics: 7th international conference, icsr 2015, paris, france, october 26-30, 2015. chapter A Case Study of Robot Interaction Among Individuals with Profound and Multiple Learning Disabilities, pages 613–622. Springer International Publishing, Cham

International Conference

8. Shukla, J., Barreda-Ángeles, M., Oliver, J., and Puig, D. (2017a). Effectiveness of socially assistive robotics during cognitive stimulation interventions: Impact on caregivers. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 62–67

Workshops

9. Shukla, J. (2017). Employing socially assistive robotics to empower cognitive stimulation. In *5th JIPI: Jornada d'Investigadors Predoctorals Interdisciplinària*, page 6, Barcelona, Spain. Universitat de Barcelona

10. Shukla, J. (2015). Using humanoid robots to convey rehabilitation therapies to disabled people. In *2nd URV Doctoral Workshop in Computer Science and Mathematics*, pages 63–65, Tarragona, Spain. Universitat Rovira i Virgili. ISBN: 978-84-8424-399-1

Chapter 2

Background

Summary. Exploration of robot-assisted interventions for cognitive rehabilitation is highly interdisciplinary work. Understanding the underlying physiological principles as well as state-of-the-art in robotics technologies is required to make significant advancements in the field. To provide a better understanding of the undertaken research, this chapter presents a short introduction of the following topics

- Socially Assistive Robotics (SAR) and its contributions to health-care
- Cognitive Rehabilitation and its benefits
- Emotion Recognition and associated machine learning techniques

Robot-assisted interventions for cognitive rehabilitation is interdisciplinary research and requires both knowledge of the psychological principles of cognitive rehabilitation and expertise in engineering models. With this in mind, we introduce the fundamentals of cognitive rehabilitation, socially assistive robotics (SAR) and emotion recognition in this chapter. While emotion recognition is a very broad field but we focus on online recognition of user's emotions during robot-assisted cognitive rehabilitation which limits the scope and makes it easier to pursue. To begin with, a review of SAR along with its definition is presented in section 2.1. We then introduce the important underlying concept of cognitive rehabilitation and its impact among individuals with wide range of mental health concerns in section 2.2. Finally, the definitions of emotion and affect and use of machine learning methods for emotion recognition are briefly summarized in section 2.3.

2.1 Socially Assistive Robotics

Care of individuals with cognitive and social disabilities is a rapidly growing area of research in SAR. Current studies in this line are primarily focused on diagnosis and rehabilitation treatment of cognitive and social disabilities (Cabibihan et al., 2013; Pino et al., 2015; Rabbitt et al., 2015). NAO[1], PARO[2], AIBO[3], Probo[4] are among several other robots that are being used in this category of research (Fig. 2.1).

Normally, SAR has been used in mental health-care in three primary roles: as a companion, as a coach, and as a play partner (Pino et al., 2015; Rabbitt et al., 2015).


(a) NAO


(b) Paro


(c) AIBO


(d) Probo

Fig. 2.1 Examples of robots used in SAR

Studies on the use of robots as companions for users have reported several benefits. For instance, reductions in physiological stress levels measured through the observation of physiological signals have been found after interactions with a social assistance robot (Kanamori et al., 2003). A study including a sample of older adults with dementia interacting with a social robot in attention and memory tasks is presented in (Tapus et al., 2009a). The results presented by the authors showed that the robot was able to effectively sustain the participant's attention in the activity. Similarly, in (Moyle et al., 2014), a companion robot called Giraff is used to facilitate the communication between users with dementia and their family. The results indicated positive reactions of the users due to the continuous communication with their family through the telepresence robot. This is another useful application of robots to enhance the engagement between the family and the person with dementia living in a long-term care.

In (Standen et al., 2014), a humanoid robot was programmed as a coach to enable teachers to achieve certain previously identified learning objectives for eleven people with ID. Results have shown that the attention during the sessions involving the robots was higher in comparison to the session in the classroom. When the robot is used as a play partner, the response of the users is positive. For instance, after several months of interaction with

[1]Aldebaran, https://www.aldebaran.com/en
[2]PARO Robots Inc., http://www.parorobots.com/
[3]SONY Corporation, http://www.sony-aibo.com/
[4]Vrije Universiteit, http://probo.vub.ac.be/Probo

a simple robot, young children with developmental disorders displayed increased social engagement (Kozima et al., 2007). Likewise, elderly users have shown a reduction in their level of loneliness when robotic pets are used (Banks et al., 2008). It important to realize that the feeling of loneliness is a common problem in users who stay for a long time in long-term care facilities, which can lead to depression in the users. In another recently published work, a Five N=1 robot-based animal assisted therapy (AAT) study was performed in adults (59-70 years) with moderate to severe ID (Wagemaker et al., 2017). In this study, during a control phase of four weeks, participants could interact with a plush seal, which was replaced by the robot seal Paro during an equally long treatment phase. The researchers concluded that robot-based AAT does not have clear beneficial effects on alertness and mood in adults with moderate to severe ID, but that positive interactions with the robot seal could be of therapeutic value in itself. Prior to this, another study described an increase in activity in a child with severe ID, as well as improvements in communication in a twin with Angelmann syndrome after interacting with Paro (Marti et al., 2005).

Preceding literature review indicates that several studies have found many positive implications in the use of social robots in therapy, even among individuals with ID but studies in this area are still limited. To point out, the robot integration in the therapies with SAR must be performed according to the indications given from the experience of health-care professionals and caregivers who know better the requirements of an effective therapy session (Cabibihan et al., 2013; Diehl et al., 2012; Pino et al., 2015; Rabbitt et al., 2015). Also, the impact of such interventions on users and the caregivers is required to be examined. This research focuses on robot-assisted interventions for cognitive rehabilitation. Related to this, the success of such interventions are not only dependent upon whether the robot successfully performs tasks, but also on the active participation from the user's part. It is important to realize that the users with mental health concerns need encouragement and emotional adaptation for an active participation. It has been shown that a robot with adaptive behavior can improve the user's task performance in the cognitive game (Tapus et al., 2009b). Accordingly, the robot behavior adaptation must be based on the patient's level of disability and their current level of emotional state and engagement. The robot needs to know the current level of emotional state and engagement of the users during the online interaction. Accordingly, this research focuses on online recognition of the user's current emotional state.

2.2 Cognitive Rehabilitation

Cognitive rehabilitation has been defined differently by various practitioners. In general, cognitive rehabilitation refers to non pharmacological and nonsurgical interventions by health-care providers to improve or restore problem solving capabilities of the brain (Halligan and Wade, 2005). It involves guided practice on a range of standard tasks related to one or more cognitive domains (Worthington, 2005). In different approaches, cognitive rehabilitation is used not only for treating acquired (through brain damage) cognitive disorders but is also applied to developmental disorders of cognition (Halligan and Wade, 2005). When applied to the rehabilitation of developmental cognitive impairments, it can benefit cognition of individuals struggling with a wide range of mental health concerns, including adults with intellectual developmental disorders (IDD) (Peñaloza-Salazar et al., 2015) and people with neurodegenerative disorders such as Alzheimer's disease (Worthington, 2005).

Among several cognitive rehabilitation approaches used, cognitive stimulation therapy (CST) is an evidence-based psychological or psychosocial intervention consisting of structured sessions of stimulating activities in a group setting or for individuals (Spector et al., 2011, 2010). The cognitive benefits of CST are well established and the results are seen as being emotionally positive and most participants reported cognitive benefits (Spector et al., 2011). Such gain in cognition can increase autonomy among these individuals which in turn can improve quality of life and hence well-being of these individuals (Worthington, 2005). At the present time, CST is being offered primarily in residential care facilities to individuals by caregivers. But it requires sustained efforts, motivation, ongoing repetition and practice (Halligan and Wade, 2005). Hence, these intervention activities demand huge efforts. Along with growing number of individuals with mental health concerns and predicted shortage of human shortage in health-care, this leads to an inability to provide cognitive rehabilitation to everyone in need.

Recent advances in SAR offers a promising hope that SAR can empower cognitive rehabilitation to make it accessible to individuals with wide range of mental health concerns in low-cost manner. But to achieve successful cognitive rehabilitation with SAR, it is important to understand the general principles of cognitive rehabilitation among individuals with mental health concerns. Moreover, critical elements for successful rehabilitation planning, implementation and evaluation around SAR is required to be investigated. It is to be noted that the differences between the words *cognitive rehabilitation*, *cognitive stimulation* and *cognitive training* are subtle and hence use of any of the above terms throughout the thesis mean the same thing which is, "non pharmacological and nonsurgical intervention activities by caregivers to improve cognitive functioning of the individual".

2.3 Emotion Recognition

Knowledge about the emotional state of a user is an important factor to provide an empathetic robot interaction experience to the individuals with intellectual disability. Understanding the user's emotion will not only lead to the correct interpretation of their actions as well as communication intents, but it can also serve as an objective assessment tool for evaluating the user's experience in any involved settings. Emotions can also be included for improving the interaction experience of these users with the robots, which can be conceived on several levels. Robots that understand the user's emotion can react in a more appropriate manner. For example, in design of the robot-assisted cognitive training activities, in which emotional states can be used to provide motivation, hints and clarifications, or can be similarly used to adjust parameters for difficulty.

There is no universal definition of emotion. Several definitions of emotion exist in context-specific way and are still argued over. As quoted by Smith and Lazarus, emotions are "differently described and explained by different psychologists, but all agree that it is a complex state of the organism", usually including strong feeling and action tendencies (Lazarus, 1991).

Theoretical models of emotions have been grounded in two competing perspectives. On the one hand, discrete models of emotions claim the existence of a number of distinct basic emotional states, such us joy, sadness, anger, fear, surprise, or disgust, the combination of which characterizes the human emotional experience (Ekman, 1992). By contrast, bi-factorial models describe all the existing emotions as a function of two core factors: hedonic valence (whether the emotion is positive or negative) and arousal (i.e. the level of excitement) (Russell, 1980).

Measures of Emotion

The presence of emotional states is evidenced through three different types of manifestations: self-report from the subject (collected, for instance, through questionnaires or interviews), changes in physiological aspects in the subject's body (e.g., heart rate or conductivity of skin), and directly observable behaviors (e.g. face expressions or body movements) (Mauss and Robinson, 2009). But the individuals with ID usually have limited ability in recognition and expression of their emotional state, hence the analysis of physiological and behavioral correlation of emotions emerges as the most useful method for monitoring their emotions (Shukla et al., 2016a).

In particular, emotion recognition from physiological signals is expedient, since it taps the pure, unaltered emotion in contrast to modalities like facial expressions which can

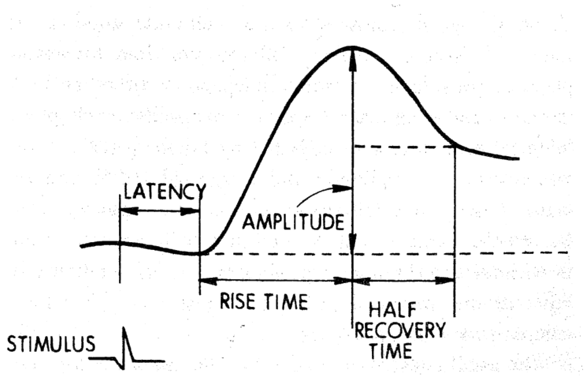


Fig. 2.2 EDA Waveform (taken from (Dawson et al., 2007)

be faked (Jenke, 2015). It also does not require the user's attention, which is important especially when assessing emotions among the individuals with ID who can be attention deficit. Recent advancements in the wearable technologies has shown a vital potential for hassle free acquisition of the physiological signals in an unintrusive manner and have thus inspired us to investigate online emotion recognition using physiological signals among individuals with ID.

EDA for Indicating Human Cognitive States

Physiological signals cover both brain signals, such as Electroencephalogram (EEG) signals and peripheral physiological signals of the autonomous nervous system (ANS), which include heart activity, respiration, temperature, Electro-dermal activity (EDA) etc (Jenke, 2015). Due to its low-cost and easy-to-collect nature, EDA measurement has been commonly used in research in psychology (Dawson et al., 2007) and as a tool for the assessment of user's experience in a variety of contexts such as recreational and serious games (Drachen et al., 2010) (Nacke et al., 2010) , driving (Healey and Picard, 2005), or patient-robot interaction (Swangnetr and Kaber, 2013). Hence, we have adopted emotion recognition using EDA signals.

EDA signal is a non-stationary signal, with mean levels usually ranging between 2 and 20 μS, and varying within a range between 1 and 3 μS for an individual. Its values typically show a slow decrease over time when the subject is at rest, and increase more rapidly when novel stimulation is introduced, and, once the stimulation is over, gradually decrease again (Dawson et al., 2007). Figure 2.2 presents a typical EDA waveform and its parameters. EDA signal is considered to have two components: a tonic, or general, level, and a phasic component, characterized by more rapid and momentary changes in EDA levels. Such momentary changes, usually associated to the presence of arousing stimuli, are called Skin

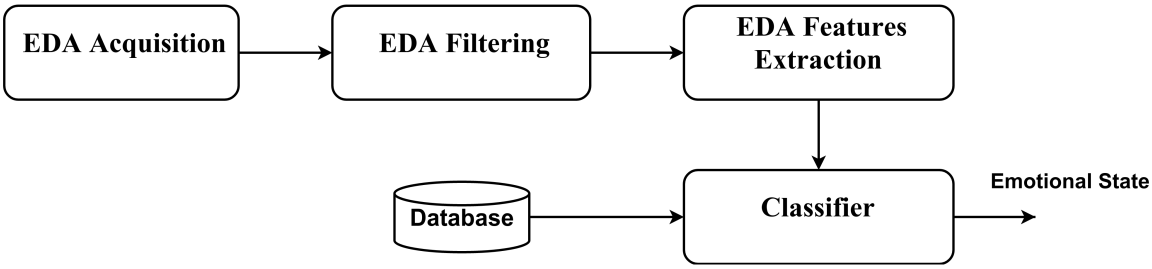


Fig. 2.3 Emotion recognition using machine learning

Conductance Responses (SCRs). SCRs can be seen as peaks over imposed to the tidal drifts in the general EDA levels.

Important Issues

The common processing pipeline for emotion recognition using machine learning is visualized in Figure 2.3 and usually includes data acquisition (DAQ), filtering, feature extraction, classification and evaluation.

For the recording of physiological signals, lightweight, wearable and wireless commercial sensors are readily available that allow real-time data collection, display, and storage. Sample wearable devices that can be used for recording EDA signals are E4 wristband[5] and Shimmer3 GSR+[6].

Use of the machine learning for the emotion recognition has been investigated in laboratory conditions since several years. However, compared to laboratory studies, our research involves the online collection and analysis of EDA signals in real-life context. This introduces several new challenges which is the primary research focus of our work and are described below:

- Database: Emotion recognition among individuals with ID has been challenged by the absence of any annotated multimodal database for them. Besides, existing databases employ non-personalized stimulation circumstances to stimulate emotions among subjects, which do not correspond to the natural real-life scenarios.

- Filtering: Compared to laboratory studies, our research introduces new challenges related to filtering of the signal. For example, motion and daily activities affect the physiological measurements and can deteriorate their quality for emotion recognition.

[5]https://www.empatica.com/e4-wristband
[6]http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

- Feature Extraction and Selection: A major limitation is that no systematic comparison of EDA features exists. Hence, a major challenge in our research to review the existing feature extraction methods for emotion recognition from EDA and identify the significant features of EDA that can be exploited in real time.

It is important to address the above issues in systematic ways to achieve emotion recognition among individuals with ID in real world settings during robot-assisted interventions.

Part I

SAR Empowered CST: Design, Implementation and Evaluation

Chapter 3

Design of Robot-Assisted Interventions

Summary. To ensure the usability of the robot-assisted interventions for cognitive rehabilitation, it is necessary to collect the requirements from the intended users. In order to achieve this objective, multiple interviews and a round table discussion was done with the experts and the caregivers working with Individuals with ID. In this chapter, we discuss the needs of the users and their caregivers, the feasibility of robot interventions, possible usage scenarios, as well as some general considerations about the robot behavior. Based upon these results we have provided a general framework incorporating the robot-assisted interventions for cognitive rehabilitation.

Our work focuses on collaborating with the intended users and their caregivers from the beginning, rather than designing a system for them. Hence, a participatory design approach was followed. The objective of the interviews was to identify *what sort of assistance can be offered to individuals with ID using SAR*. To achieve this objective, multiple interviews and a round table discussion was done with the experts and the caregivers working with Individuals with ID.

3.1 Attendees

Seven experts were interviewed and their profiles are given in table 3.1. The participation of experts was voluntary and they were not remunerated. As can be observed from table 3.1, experts involved in the interview had an extensive work experience (an average of 21 years total experience and 12.43 years of experience in assistive technologies) among individuals with ID. These experts were chosen based upon their large experience of using assistive technologies such as Basal Stimulation (Bienstein and Fröhlich, 2003), Tactile Gaming Consoles (Peñaloza-Salazar et al., 2015), etc., among individuals with ID. Experts at the Ave Maria Foundation were even familiar with the use of robots among individuals with ID.

Due to their vast experience in using assistive technologies among different age group of individuals with ID, these experts possessed a strong background to discuss the question at hand.

Table 3.1 Expert Profiles.

Expert ID	**Education**	**Experience (Years)**	
		Assistive Technologies	Total
E01	Post Graduation in Mental Health and Behavioral Disorders in Persons with ID	10	17
E02	Graduate in Child Psychopathology of Adolescents and Master in Clinical Mental Health	10	17
E03	Master-Logopedia: Language and Speech disorders	10	19
	Post-Graduate-Knowledge and Care in Intellectual Disability		
E04	Postgraduate specialization in pediatric physiotherapy	10	27
E05	Postgraduate specialization in Occupational Therapy	21	23
E06	Postgraduate specialization in Therapeutic Pedagogy	16	21
E07	Diploma in basic general education teachers. Special Education specialty.	10	23

3.2 Methodology

Each expert was interviewed individually during the first stage and the interviews lasted between 45-60 minutes. Experts were presented with certain open-ended questions and further discussions continued around answers given by the experts. Following are the main questions that were presented during the interviews:

- What are the different types of activities that are required to be performed among individuals with ID? Can the activities be performed by a robot? If yes, then how?
- What are the different metrics (e.g., engagement rate, etc.) that can be used for evaluation of such a performance?
- Is there anything that you can think of, which a robot can do better than you during an interaction?

Once the individual interviews were over, a round table discussion was conducted with experts during the second stage to define a suitable use case of SAR. Prior to design of the use case, following constraints were kept in mind:

- A direct, face-to-face, interaction scenario is considered where a robot is interacting on a one-to-one basis with individuals with ID.
- Robot is equipped with the state-of-the-art voice interaction and visual perception abilities.
- Individuals from the chosen sample have basic perception and interaction abilities.

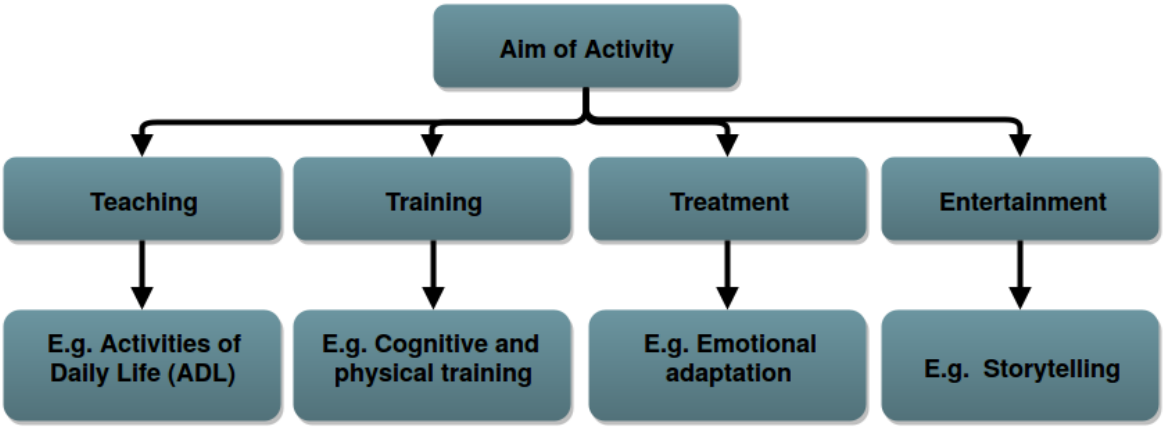


Fig. 3.1 Types of Robot Interaction Activities

3.3 Results

3.3.1 Types of Activity

Experts gave different examples of activities that are required to maintain/or exercise the cognitive ability of the users. To point out, experts mentioned that an interactive activity is conducted with the participant to serve a specific stimulation purpose and it varies according to the intellectual and physical ability of the participant. Altogether, following are some examples of the activities mentioned by the experts:

- Activities are designed to provide knowledge to the user about daily routines, such as brushing the teeth, taking a shower, etc.
- Activities are designed to encourage users to perform physical exercises, such as by dancing, by mimicking actions, etc.
- Activities are designed to provide knowledge and awareness about their body.

Based upon the above examples by the experts, four primary reasons for performing an activity were identified, which are illustrated in figure 3.1. Finally, the description of each category is summarized in table 3.2.

3.3.2 Advantages of Robot Assistance:

A robot can offer certain unique advantages over human caregiver during the above activities. Some suggestions, as put forward by the experts, are as follows:

- By using external sensors, the robot can measure and analyze the physical and/or mental state of the users in an objective manner.

Table 3.2 Objectives

Objective	**Description**	**Role of the Robot**	**Sample**
Teaching	This category aims to teach a specific behavior or skill to the users via an interactive activity. Examples can include *activities of daily living (ADL)* such as brushing teeth.	Teacher	The robot can employ real/fake objects (such as toothbrush, toothpaste, etc.) and/or printed images to demonstrate and explain the sequence of brushing teeth through conversation.
Training	Aim is to provide a training to the users. Examples can include physical training, exercising motor functions or cognitive rehabilitation aiming to stimulate different cognitive capabilities of the users.	Coach/ Instructor	The robot can direct the physical training of the user. Initially, the robot shows the motions that the user must imitate. The robot can monitor the motions imitated by the user and if required, the robot again demonstrates the motion and encourages the user to imitate the similar motion taking into account the user's motor limitations.
Treatment	This category aims to provide a treatment to the users at a cognitive level. An example of this category could be of *Emotional Adaptation (EA)*, where the user is not feeling *good* and interactive activities are used to encourage his/her mood.	Companion	The robot can monitor the emotional state of the user. Upon detecting a negative emotional state among the user, the robot can play the user's favorite song to make the user's emotional state positive.
Entertainment	Activities that aim to entertain users without specific focus on any of the above categories, can be classified in this category. An example of this category includes storytelling.	Play Partner	The robot can tell a story while demonstrating the actions occurring in the story such as bending, supporting, turning, etc. to the user, encouraging them to action it out. The robot can employ optional equipments to suit the theme; e.g., hula hoops representing puddles, stones representing rocks to climb over.

- Robots are not affected by any emotions. Consequently, robots can work for longer periods in comparison to humans, especially when the several repetitions of the activity is required.

- Robots can utilize Baby talk style, entertaining talking styles such as employed by popular mobile application Talking Tom[1] and can provide a home cinema like experience with coordinated control of additional devices and/or sensors during the activity.

3.3.3 Role of the Robot

Once the types of the activity required for the users were identified, experts envisioned a sample role of a robot during a particular activity. Experts agreed that to serve a meaningful usage, the robot will have to help achieve the objectives of a specific activity. Hence, for each of the previously identified four activity types, four roles of the robot were determined and are detailed in table 3.2.

3.3.4 Intervention Strategies

Experts pointed out that a particular role-playing can be realized by employing different intervention strategies. In order to realize a particular role in the activity, either one or a mix of the following intervention strategies was analyzed for the robot:

- Turn-taking

- Imitation

- Feedback

As an illustration, an activity of dancing to entertain can be designed by the following different ways using any one of the intervention strategies:

- Turn-taking: Robot makes a step of the choreography and the user makes the next step of the choreography. In this way, both take turns and the activity progresses to entertain the user.

- Imitation: The robot takes a step according to the planned choreography and the user mimics the same step. This way, the user gets entertained by imitating the robot's dance steps.

[1]https://itunes.apple.com/us/app/my-talking-tom/id657500465?mt=8

- Feedback: User dances and the robot provides a positive feedback to encourage him/her.

Depending upon the user's preference and abilities, any one of the above mentioned intervention strategies can be employed for creating an activity. In addition, selection of an intervention strategy also depends on the personal preferences of the caregiver who designs the activity. Sometimes, a mix of two or all of the three strategies can be used to achieve an activity design, where the robot and the user can be involved simultaneously in turn-taking, imitation and/or feedback.

3.3.5 Metrics for Evaluation

The robotic intervention activity must be evaluated to know its effectiveness. However, the metrics are not standard for any such evaluations. Following are some of the methods suggested by the experts during the interviews:

1. The outcome of the activity as compared to the aim of the activity; e.g, if the aim of the activity was to teach the user about taking a shower, did the user become aware of taking a shower?

2. Questionnaire about the behavior of the user during the intervention.

3. Pleasure or comfort of the user during an activity as determined by monitoring their emotional state and their engagement during the activity.

The experts agreed that both the first and the second method of evaluation while assessing the impact on the users, must be performed with the help of the caregivers, as users can not do the evaluation for themselves due to their limited intellectual abilities. But, the evaluation using the third metric can be tricky among individuals with ID due to the limited ability in recognition and expression of their emotional state. Hence, experts were further asked to define the engagement and emotional state as observed by them among these individuals. While experts encounter the engagement of participants during an activity on a day to day basis, but were troubled most to put in words what they mean by engagement during an activity. Following are some of their answers:

- *If participant is **doing something** related to the activity, even if there is no progress in the activity* - E03.

- *If participants are focused on the activity, then they are engaged* - E05.

Based upon above opinions, following definition of engagement was adapted for evaluation of activities performance, where engagement is defined as "the process by which individuals involved in an interaction start, maintain and end their perceived connection to one another" (Sidner et al., 2005).

It is to be noted that above evaluation metrics are focusing on impact of the robotics interventions on the users only, so experts were further asked to define metrics for the impact on the caregivers during the robotic interventions. Moreover, all experts agreed that execution of cognitive stimulation interventions represents immense burdens on caregivers in terms of time and labor costs. Since any reduction in such type of burden on caregivers can be directly utilized towards the individual attention of the users, following evaluation metrics were determined to evaluate the impact on the caregivers:

- Reduction in workload of caregivers by the use of robots.
- Increase in personalized attention towards users due to above reduction.

The experts pointed out that if having a robot during the interventions can reduce the workload on the caregivers and can allow them to devote more time for the personalized attention of the users, it will certainly be a measure of success.

3.3.6 Proposed Robotics Assisted Interventions

Results presented above clearly indicates positive potential of robotics assisted interventions on the individuals with ID and their caregivers. However, two major shortcomings were observed:

- Users need encouragement and emotional adaptation for an active participation. It has been shown that a robot with adaptive behavior can improve the user's task performance in the cognitive game (Tapus et al., 2009b). Moreover, the experts recommended during the interviews that the robot behavior adaptation must be based on the patient's level of disability and their current level of emotional state and engagement.
- Emotion monitoring among these users becomes very difficult task due to the limited ability in recognition and expression of their emotional state. It also makes the engagement evaluation very trivial and hence the engagement status as reported by caregivers can be very subjective.

Owing to the limited ability in recognition and expression of their emotional state, the analysis of physiological and behavioral correlation of emotions emerges as the most

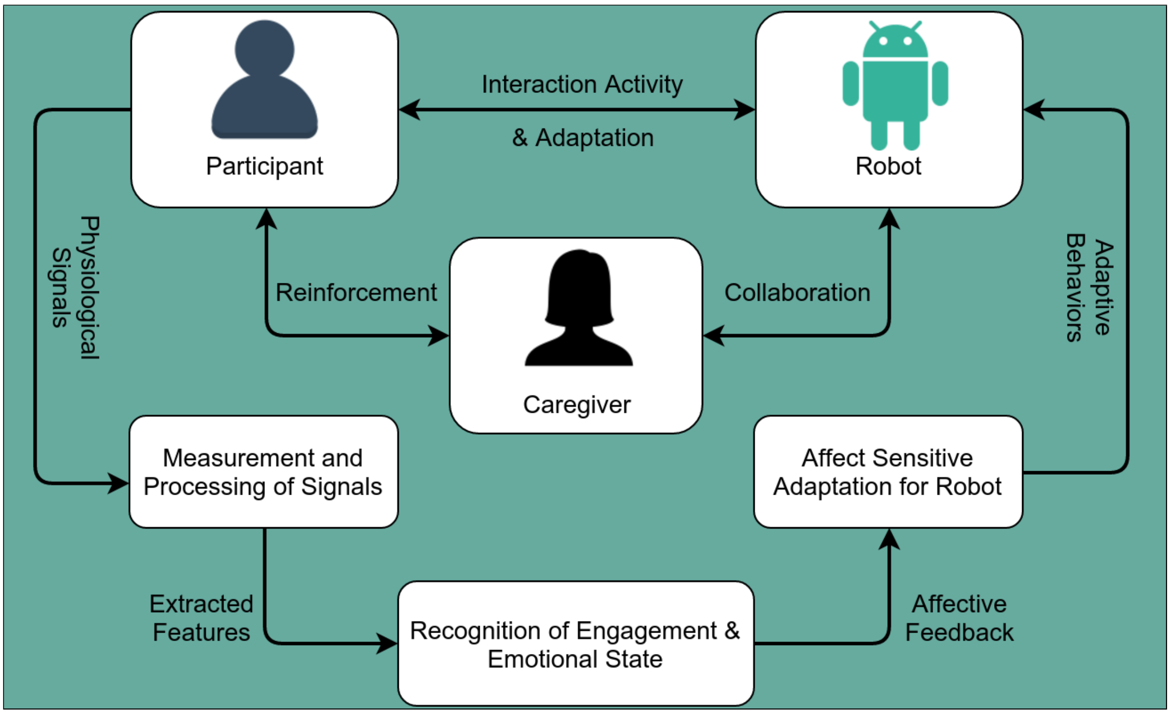


Fig. 3.2 Feedback based SAR

useful method for monitoring emotions among individuals with ID (Shukla et al., 2016a). Consequently, Identifying the critical need of emotional monitoring among these users, we present a general framework of the robot-assisted interventions including feedback of involved participants in figure 3.2. The framework in figure 3.2 can be understood as follows:

1. Based upon the user's requirements and needs firstly the caregiver identifies a specific stimulation activity in one of the categories as mentioned in figure 3.1. Then, the caregiver designs an activity in the identified category. Consequently, the robot collaborates with the caregiver during the execution of the designed activity with the participant.

2. During the activity, the participants are equipped with different physiological sensors to obtain biosignals such as Heart Rate (HR), Electro-dermal Activity (EDA), Electromyography (EMG), Electroencephalogram (EEG), etc. High resolution video cameras can also be employed to obtain the audiovisual recordings to interpret vocal and facial expressions. In addition, RGB-D and/or infrared cameras are fitted to analyze eye movements, body posture and head movements of participants during the above interaction with the robot.

3. Various signal processing techniques are applied online to extract meaningful features from the above signals and camera data.

4. Automatic emotional state classification and engagement prediction is done from above features using assorted machine learning and data mining algorithms.

5. The above information is fed to the robot, based upon which the robot plans a new course of action that can be either to continue the activity or to encourage the participant or to halt the activity.

6. The robot takes the newly planned actions based upon the feedback obtained from the participant in collaboration and in agreement with the caregiver. The specificity of the communication between the robot and the caregiver can be determined based upon the technical abilities of the robot.

Some comments regarding the use of biosignals are necessary to mention. Firstly, the sensors used to obtain the physiological signals in real-time must be hassle free for the participants and should cause a minimum level of distraction, as the targeted users are individuals with ID, who get easily distracted by any device or sensor attached to their body or placed around them. Any distraction caused by these sensors will lead to adverse effects on the productivity of interaction activities. Sample wearable devices that can be used fitting these restrictions are E4 wristband[2] for recording EDA and HR data, Emotiv Epoc+[3] for recording EEG data etc. Secondly, combining different modalities helps to obtain more information. Hence, the information from biosignals can be complemented by the camera data to obtain more meaningful knowledge about participant's behavior during the interaction.

Encouragement and emotional adaptation methods as suggested by the experts, depend upon many things e.g. foreknowledge of the participant's likes/dislikes, interests, interest of the educator, etc. Some examples of encouragement actions as stated by experts are: playing a music or video of user's choice, allowing the user to interact with his dear ones, e.g, by making a call to a family member, replacing the current activity with another one that is more interesting to the user, etc.

3.3.7 Use Case

Based upon the requirements and abilities of these individuals, the following use case was agreed upon by the experts:

[2]https://www.empatica.com/e4-wristband
[3]https://www.emotiv.com/epoc/

1. A user enters the room, in which robot is kept.
2. The robot recognizes the user.
3. The robot initiates a warm up conversation and invites the user to take a seat.
4. If user's response is not positive, the robot tries to make the user feel comfortable by encouraging him/her.
5. After the user takes the seat, the robot proposes one activity prescribed for the user.
6. If the user agrees to play, the robot starts the execution by explaining the activity to the user.
7. If the user does not understand the activity, the caregiver intervenes to assist the user in understanding the activity.
8. If the user understands the activity, the robot starts the execution.
9. The robot constantly monitors the engagement, emotional state and progress of the activity and at the end of the session saves all the data for post activity evaluation.
10. If the user is not engaged or distracted, the robot encourages the user to perform the activity.
11. If the user's emotional state is not positive, the robot applies behavior intervention plans that encourage the user to perform the activity.
12. If the progress of the activity is slow as per the expectations of the user's ability, then the robot helps the user by providing assistance with the performing of the activity.
13. If the activity shows no advancement as per the user's ability or the user does not want to continue, the robot evaluates the situation and determines if it must restart the activity again or definitely finalize its execution.
14. At the successful completion of an activity, the robot saves all the useful data in a detailed report for post activity evaluation.

The use case is illustrated in figure 3.3.

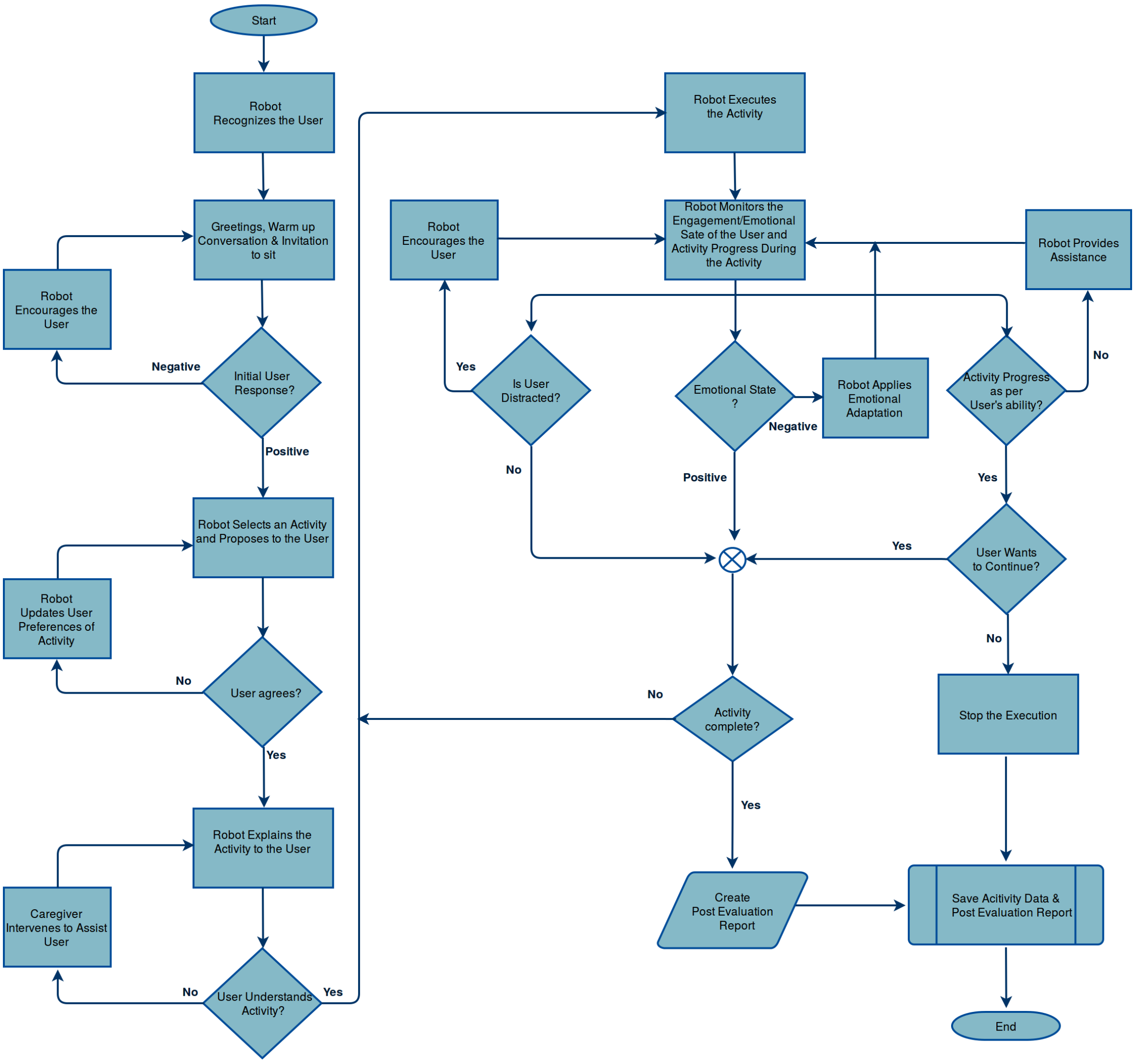


Fig. 3.3 Use Case

3.4 Conclusion

As can be observed via section 2.2, previous researches have reported positive effects of robotic interactions in mental health care interventions. To summarize, responses obtained from the interviews suggest that robotic interactions among individuals with ID can have similar positive effects on cognitive care of individuals with ID.

Different types of interaction activities designed and presented in table 3.2 clearly suggest that robotic interactions among these users can provide useful benefits in different categories while serving specific stimulation purpose. While performing these activities, the robot will have to assume different roles which is also depicted and explained in table 3.2. Similar classification of robot roles in mental health care interventions have also been identified by (Rabbitt et al., 2015). Additionally, different intervention strategies for role-playing by the robot as identified in section 3.3.4 can help to personalize the delivery of intervention activities to the needs, requirements and preferences of the users. Metrics of evaluation as proposed in section 3.3.5 are directly derived from the expert caregivers working in the field and hence it can serve as a measure of success in evaluating a given robotics intervention system and its impact on the user and their caregivers. Similarly, the proposed use case and the evaluation metrics take into account the requirements and benefits for both the user and their caregiver and hence can save the time in implementation of the robot assisted cognitive stimulation activities and help the caregiver accomplish his/her task of stimulating the user more effectively.

Chapter 4

Robot Assisted Interventions: Impact on Users

Summary. Motivated by the interview responses and the promising advancements about the use of interactive robots for rehabilitation, we envision that SAR can benefit cognitive rehabilitation of individuals with ID. However, a prior assessment of the impact of SAR during such interventions on the users is required. Accordingly, in this chapter, we present a case study regarding interaction of individuals with ID with a humanoid robot NAO in different possible categories of robotic interaction. Also, only few studies have shown any significant comparison in remedial therapy using interactive robots with non-robotic visual stimulations. Hence, we compared the response of robotic interactions with non-robotic stimulations caused by a tactile gaming console. The results demonstrate the positive effects of such interventions on the users, mainly an increase in the engagement. In addition, the comparison results indicate that robot embodiment is more beneficial in engaging the users over tactile gaming consoles.

Based on the possible usage scenarios and general considerations derived in chapter 3, we conducted a case study to evaluate the initial effectiveness of robot interventions for users in different intervention categories. In general, SAR claims many promising advancements about the use of interactive robots for rehabilitation of individuals with ID. However, many other assistive technologies are also available to offer assistive therapy to these individuals. Therefore, a related question is: *Why is the use of SAR encouraging for rehabilitation among individuals with ID?*. This work aims to answer the above question also by making a comparison of robot interactions with non-robotic stimulations caused by a tactile gaming console.

4.1 A Case Study of Robot Interaction

4.1.1 Methodology

Separate interactive activities were designed in previously identified four broad categories of interactions, to analyze the response of individuals with ID in these different possible categories. Accordingly, identified activities are listed in table 4.1. The activities were chosen based upon the following criteria: relevant representation of the category by the activity of choice, simplicity for participants at the execution level and ease of implementation in NAO robot.

Table 4.1 Activities

Category	**Representing Activity**	**Description**
Entertainment	Dance choreography	NAO performs a dance composition while singing a song. NAO was not interacting with the participant at any level while participants were allowed to observe and respond without any restrictions.
Training (Physical)	Touch my head	The robot asks the participants to touch its head, feet or hand. The participants are expected to respond according to its instructions.
Treatment (Emotional Adaptation)	Guess emotions	The robot tells a short story to the participants, and in between, asks questions related to the emotional state of the character in the story. The robot helps participants answer the questions, helping them to learn about different emotions.
Teaching	Learn the senses	NAO prompts the participants to present an answer image corresponding to a particular sensory activity of the human body. If the participants do not answer within a certain period, NAO encourages the participant by providing some clue about the answer image. The robot provides a positive feedback when the participant gives the right answer.

Participants

Six individuals with ID were selected to conduct this research. User's inclusion criteria was defined as follows:

1. Age equal to or above 18 years.

Table 4.2 Participants

ID	**Gender**	**Age**	**Condition**	**Disability (%)**
ID01	F	65 y, 4 m	Moderate Intellectual Disability	85
ID02	F	42 y	Severe Intellectual Disability	86
ID03	F	48 y, 8 m	Severe Intellectual Disability	87
ID04	F	33 y, 2 m	Moderate Intellectual Disability	79
ID05	F	67 y, 10 m	Moderate Intellectual Disability	86
ID06	M	44 y, 6 m	Severe Intellectual Disability	75

2. An official diagnosis of ID as done by *Assessment and Guidance Services for People with Disabilities (CAD Badal)* organization[1]. The assessment criteria applied by the CAD Badal to compute the patient's level of disability, is defined and described in the Spanish laws and are published in official state gazette[2].

3. Living in the FAM residence facilities for at least 3 years.

4. Familiarity with the NAO robot through random interactions.

5. Their guardians have provided written consent to take part in the study.

Details of the selected individuals are given in table 4.2.

Set-Up

Figure 4.1 shows a participant interaction scenario with NAO during the activities. The NAO NextGen (Model H25, V4) humanoid robot was used for this case study. The reasons for choosing NAO in this case study were as follows:

1. Its small humanoid form and extreme interactivity makes it really endearing and lovable to the type of individuals in our study.

2. NAO can move, recognize objects and people, hear and even talk to individuals. These features make it suitable for all the categories of clinical applications of interactive robots.

3. The size of the NAO is that of a human toddler, which is found to be the most suitable for clinical applications with interactive robots (Giullian et al., 2010).

4. NAO is easily and commercially available.

Procedure

For each trial, the robot was placed on a table in a position as required to initiate the desired activity. Then, the participant is brought to the trial room by the caregiver and takes a seat in front of the robot. Figure 4.1 shows a general position of the participant with the robot. During the activity, only the caregiver stays in the room with the participant while the researchers observe the whole situation from outside of the room. The care taker observes

[1] Generalitat de Catalunya, http://web.gencat.cat/ca/inici/
[2] BOE NÚM 22 del 26/01/2000, https://www.boe.es/boe/dias/2000/01/26/pdfs/A03317-03410.pdf

Fig. 4.1 A participant interacting with the robot

the participant during all the activity and does not initiate any communication on its own but responds to participants. Duration of the trials vary between 15-30 minutes depending upon the activity and the participants, while actual robot interaction during trial lasted between 5-10 minutes. Finally, all six participants participated in each of the four interaction activities.

Evaluation

Motivated by the evaluation metrics presented in section 3.3.5, evaluation of the above studies was done by using the following evaluation methods:

- **Performance of the participants** against desired responses of activities. Depending upon the time duration and correct or wrong responses, performance of the participants were recorded as perfect, good, regular or no response.

- **A questionnaire** adapted from GARS-2 Montgomery et al. (2008), WHODAS 2.0 Gold (2014) and ABS-RC: 2 Nihira et al. (1993). It mainly comprised the questions to evaluate the interaction and communication intents of the users while observing their behaviors. To the author,s best knowledge, no method or scale is available for the evaluation of robotic interaction effects. Thus, the authors adapted the existing aforementioned assessment tools for evaluation purposes and to enhance the decision-making in such trials. Further, details regarding the questionnaire can be found in appendix A.

The proposed questionnaire presented above was used to evaluate the disability behavior of the participants in normal situations. Then using the same questionnaire, disability behavior during interaction with the robot was reevaluated. A difference that represents the improvement in disability behavior was then calculated using equation 4.1.

$$improvement = db_{normal} - db_{robot} \tag{4.1}$$

Where db_{normal} and db_{robot} are the disability behavior of the users as shown during their daily life interactions and during robot interaction, respectively, in similar and comparable situations. An example situation in which db_{normal} was measured is when the individual participant was interacting with another user of the care facility. These variables vary between zero and 100 percent, where zero represents non observed disability behavior in a particular situation and 100 percent indicates full disability behavior observed based upon the questionnaire.

4.1.2 Results

Table 4.3 shows performance evaluation of the participants for all three interactive activities. Each activity consisted of 3 tasks in respective categories. Thus, the total number of responses recorded from six participants was 54 (6 participants * 3 activities * 3 tasks). The evaluation of response was done by a psychiatrist taking into account the time and support used by the participant to accomplish the task.

As can be observed from table 4.3 that *among a total of 54 observations, only 2 times participants did not respond at all to the robot. Most of time the response was very positive, which can be seen by a total number of 33 perfect responses*. Some important observations for delayed or no response scenarios are summarized below:

1. Participants were attracted (absorbed) with the robot hence, were not able to concentrate on the execution of activity.

Table 4.3 Performance of participants against activities

Activity	Perfect	Good	Regular	No Response
Touch my head	14	1	2	1
Guess emotions	9	5	4	0
Learn the senses	10	3	4	1

2. Restrictions in technical abilities of the robot, (e.g. image recognition).

3. Participants were excited with the robot during activity hence, sometimes were responding before even completely listening to the robot.

Table 4.4 Engagement rate (% duration) of participants against activities.

ID	Dance	Touch my head	Guess emotions	Learn the senses
ID01	96.60	100.00	100.00	100.00
ID02	64.56	100.00	100.00	100.00
ID03	93.20	100.00	100.00	100.00
ID04	100.00	100.00	100.00	100.00
ID05	100.00	100.00	100.00	100.00
ID06	98.06	100.00	100.00	100.00

The definition of the engagement as given in section 3.3.5 was used for the evaluation which was done by the caregivers. This definition was used as all the experts agreed that it was easy to adapt for the caregivers to determine the participant's engagement in the activity. Engagement rate of participants against activities are listed in table 4.4. It indicates the percentage of time in which the user was engaged with the corresponding activity during the total time of the execution. This evaluation was done by one caregiver who was familiar to all the participants. As can be seen from table 4.4, the minimum engagement observed during the six sessions (6 users x 1 activity) of the non-interactive activity (Dance choreography activity in table 4.1) for all the six user's was 64.56%. It can also be seen from table 4.4 that the observed engagement was 100% in 18 sessions (6 users x 3 remaining activities in table 4.1) of the interactive activities with the robot. Therefore, it indicates that the interactive activities involving the robot are appealing to the users.

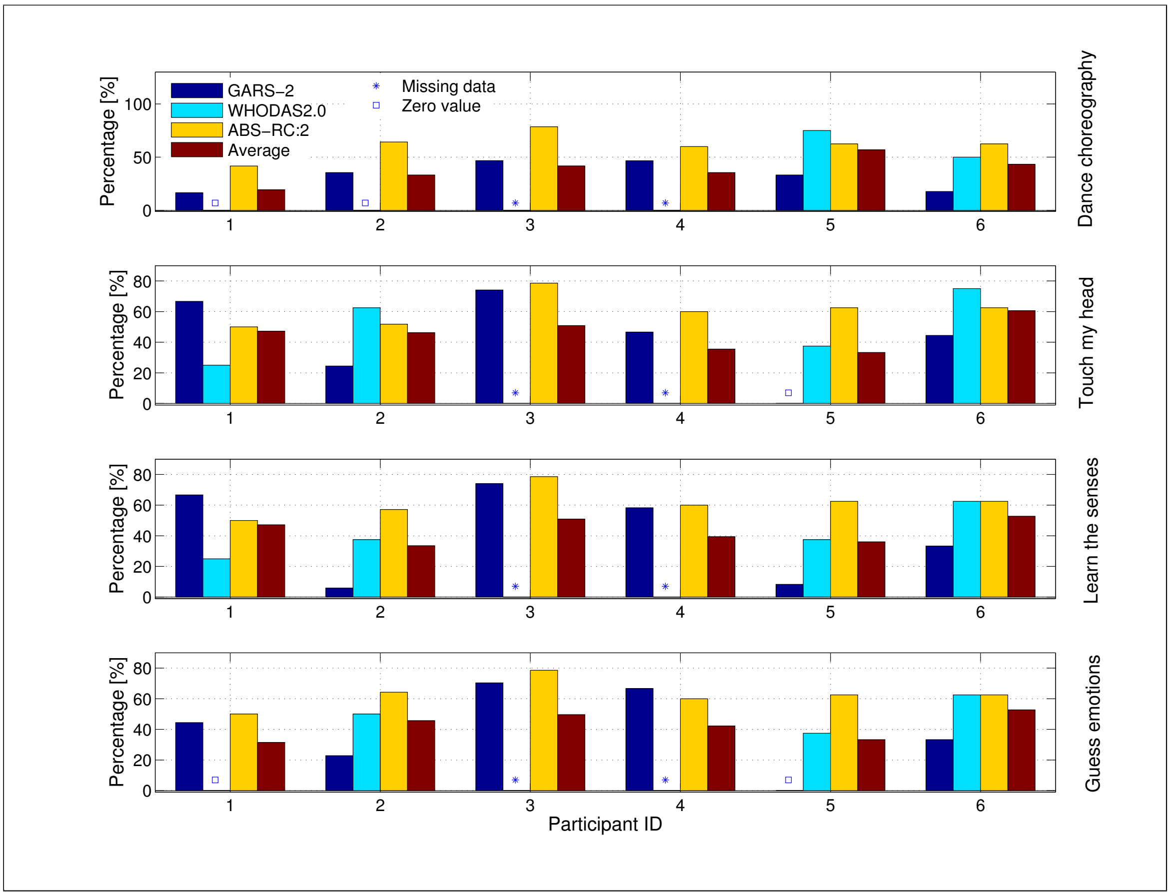


Fig. 4.2 Improvement in the observed disability behavior where *improvement* is calculated as per equation 4.1

Same caregiver carried out the evaluation task based upon the questionnaire described above and the results are presented in figure 4.2. A higher value of *improvement* in the bars shown in figure 4.2 represents more improvement in the user's response, namely a reduction in the disability behavior observed during the robot interactions as compared to the normal or basal situations. It is important to highlight that the users were familiar with the robot because they have known it for more than a year before the beginning of the experiments. Thus, the results observed in the experiments can not be due to the novelty effect of the experiment with the robot but due to the interactions with it.

Important observations can be summarized as follows:

- **Patients with different levels of ID behave in distinctive manners while interacting with the robots**. It is visible from the chart that the bar graph of all the participants are different from each other; hence, all the participants reacted distinctively to the robot in different activities. Thus, it supports the claim that *users with different levels of*

ID behave in distinctive manners while interacting with the robots. The primary reason could be that all the participants have different levels of intellectual abilities; hence, they perceive the robot and robot activity differently according to their intellectual levels and hence their response was also distinctive. Moreover, all the individuals have varying profiles (age, gender, diagnosis) which can also influence their responses. *It strongly indicates the necessity of providing customized interaction for all individuals.*

- **Patients with higher disability show lower engagement towards the robot.** The participants with ID02 and ID03 observed the lowest engagement rates as 64.56% and 93.20% respectively. These are the only participants with severe disability hence, it can be concluded that *users with higher disability show lower engagement with the robot*. The higher the ID is, the lower the interaction capabilities are. Hence, *robot activities must be designed in more attractive ways to achieve the similar effects in individuals with higher ID.*

- **A robot with interactive activity attract higher attention and improvement among users, in comparison to non-interactive activities.** During the interactive activities the engagement was always high. It indicates that the robot was able to attract full attention of the participants when performing any interactive activity with them, though the participants sometimes get distracted when they were not involved in any interaction with the robot. It supports the claim that *an interactive robot causes higher attraction among the participants*.

- **Humans are easier to follow than robots.** Some individuals in this study were not able to follow the indications given by the robot on several occasions and thus the caregiver intervened to make sure that they understood the indications given by the robot. Research suggests that for ASD affected individuals, the communication with humans is harder due to complexities involved in verbal and non-verbal communication and hence they can follow instructions from a robot in a relatively easier manner than following instructions from a human (Ricks and Colton, 2010). This observation was found true for participants ID02 and ID04, who were diagnosed with ASD as well. However, individuals affected with ID require a rich emotional interaction that can lead to an effective communication (Bellamy et al., 2010); hence, it supports the claim that *the affective communication of human beings is one of the most important reasons of why humans are easier to follow than robots for individuals with ID*.

- **Side-effects from robot interactions were unclear.** The robot interactions can generate not only a positive but a negative effect also among the participants. But it

Fig. 4.3 ARMONI Workstation

can be seen from the figure 4.2, that no negative values were obtained for any of the participants indicating that the robot interactions designed in the above fashion did not cause any negative effects on the participants. Hence, it can be stated that *side-effects from short-term robot interactions were not visible and a long term negative effects assessment is required.*

4.2 Comparative Study

SAR claims many promising advancements about the use of interactive robots for rehabilitation of individuals with ID. However, many other assistive technologies are also available to offer assistive therapy to these individuals. Therefore, now, the question that arises is: *Why is the use of SAR encouraging for rehabilitation among individuals with ID?* This work aims to answer the above question by evaluating the effectiveness of robot interactions against non-robotic visual stimulation by making a significant comparison in similar situations (Shukla et al., 2016b).

4.2.1 Methodology

This research was conducted with the same set of individuals that were involved in the previous study and whose information is listed in table 4.2. Non-robotic stimulation was provided by ARMONI (Fig. 4.3), which is a computer-based tactile gaming console, developed to rehabilitate different cognitive functions among individuals with ID (Peñaloza-Salazar et al., 2015).

To make an ideal comparison, four unique activities were identified from the available set of activities in ARMONI, representing similar categories as used in the earlier robot interaction study. The activities were chosen based on following criteria:

1. The similarity with the activities used in earlier robot interaction study.
2. Relevant representation of the category by selected activity.
3. Simplicity for participants at the execution level.
4. Availability in ARMONI.

These activities are listed in table 4.5 and their details are described as below:

1. Dance Video: This activity was aimed to observe the response of the participants towards a dance video of the robot. In this activity, a video of NAO performing a dance composition while singing a song was shown to the participants. The dance routine was exactly similar to as used in previous robot interaction study.
2. Identify the Sounds: The aim of this activity was to induce a target behavior in participants. In this activity, participants were presented with a human figure on the screen and name of the body parts are played one after another. Participants are supposed to touch the respective body parts of the figure on the tactile screen.
3. Tale Account: This activity is aimed to provide feedback and encouragement to the participant to achieve a certain target behavior. ARMONI prompts the participants to touch an answer image corresponding to a specific query related to a festival which

they are aware of. If the participant is not able to answer within a certain time period, they are encouraged by providing some clue about the answer. ARMONI provides a positive feedback on accomplishing the right answer.

4. Coordination: ARMONI works as a learning tool for the participant. A table of two columns, first column having different color paints and second empty, is displayed on the screen. Images of the same object in different colors are also shown on the bottom of the screen. Participants are required to put the matching color image in empty column of table. The activity helps participants to learn about different colors and ordering.

The evaluation was done adopting the same methods which were used to evaluate the earlier research on robot interaction study as explained in section 4.1.1.

Table 4.5 Activities

Activity ID	Activity used in ARMONI (Shukla et al., 2016b)	Activity used in Robot Interaction Study (Shukla et al., 2015)
1	Dance Video	Dance choreography
2	Identify the Sounds	Touch my head
3	Tale Account	Learn the senses
4	Coordination	Guess emotions

4.2.2 Results

This subsection presents the results (Fig. 4.4) obtained by comparing robot interaction effects with tactile gaming console stimulations.

The outcomes of the study and their investigations are as follows:

- **Physical robots have more engagement capabilities than their videos.** By comparing the engagement results during first activity, it was found that all the participants

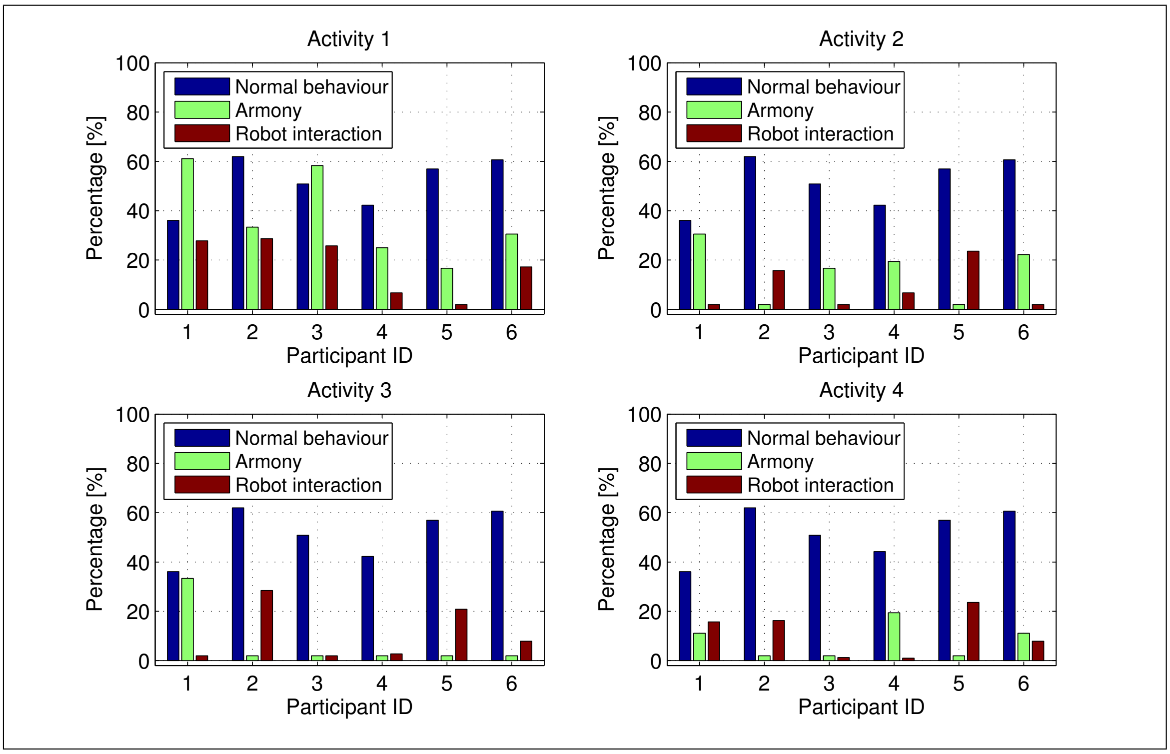


Fig. 4.4 Average disability percentage observed for each user in the different scenarios.

were more engaged with NAO dancing in front of them in comparison to the video of the same NAO choreographed and presented to them using ARMONI. It strongly supports the claim that *physical presence of a robot is more engaging than the corresponding video*. It can be explained by their ability to perceive the robot in a 3D environment allowing higher engagement.

- **Tactile gaming consoles can be used complementarily with robot interactions for an effective therapy in more users.** ARMONI was found to be more effective than robot interactions in some activities among participants. The effectiveness has been estimated from the response observed in the participants to both therapies (using ARMONI or using the robot interactions) with respect to its response in the same basal behavior. *It indicates that a tactile gaming console such as ARMONI can serve as a complementary tool in therapeutic benefit of participants.*

4.3 Conclusion

From the interactive sessions between the robot and the users in the different activities, it is evident that a robot-assisted intervention can be very helpful in drawing the user's attention, keeping them engaged in the activity. Furthermore, it is observed that *customized interventions* must be used to catch and keep the patient's engagement during the whole intervention through the diverse variety of sensory stimulus available in a robotic platform. For example, the "a priori" knowledge of the user's preferences (favorite song, favorite color, etc.) and his/her chronological evolution in the previous sessions can be used to adjust a customized, engaged and entertaining effective intervention for the user.

The first case study suggests that there is a need for a robot that is more attentive to the patient's response and his/her engagement to the therapy, by continuously tracking patient's interaction with the robot based intervention system. Therefore, real-time feedback from the patient's response must be incorporated into the system in order to assure the development of effective robot based interventions. Furthermore, the continuous improvement of the algorithms with the robot-based interventions developed taking advantage of the long expertise of caregivers and health professionals could have a positive benefit in places with a reduced number of professionals where an individual and customized intervention is not always possible. More advanced interaction functionalities using real-time feedback information will lead to the development of human-like interactive sessions between the robot and the user which will directly lead to benefit the patients.

Detailed analysis of the comparison results argued about the utility of robots over monitors and have indicated that robot embodiment is more beneficial in engaging the users over simple tablets or screens but at the same time, they serve better as a complementary rather than an alternative tool.

Chapter 5

Robot Assisted Interventions: Impact on Caregivers

Summary. Execution of cognitive stimulation interventions for cognitive training of individuals in need represents a significant burden on caregivers in time and labor costs. Current research evaluates the effectiveness of the SAR empowered cognitive training activity of *Bingo Musical* among thirty individuals with ID in multi-center trials. The results of the research confirm a significant reduction in caregivers burden and raise a concern about the need for a specific training of the caregivers to take maximum advantage of SAR in health care.

Execution of cognitive stimulation interventions for cognitive training of individuals in need represents a significant burden on caregivers in time and labor costs. Recent advancements in Socially Assistive Robotics (SAR) research can be exploited to reduce caregivers burden by work sharing with robots and supplementing/complementing human resources in the execution of interventions. The aim of the present research was to evaluate the impact of the robotics-assisted interventions on the caregivers of individuals with ID at residential care facilities. For this purpose, an activity named *Bingo Musical* was used, which falls in the *Entertainment* category of the classification scheme presented in figure 3.1. It is to be noted that since primary aim of this research was to evaluate the impact on the caregivers and not on the users, only one of the four categories presented in figure 3.1 was used. The focus of this work was to do a multi-center trial with multiple caregivers.

5.1 Methodology

A series of cognitive stimulation sessions among individuals with ID were conducted. There were five groups of users and each group of users participated in two sessions, conducted in different days, one with the robot and the other without the robot. The sessions were conducted at the following three residential and clinical facilities:

1. *Ave Maria Foundation (FAM)*[1]
2. *L'Espiga Foundation*[2]
3. *ASPACE Confederation*[3]

Two caregivers at FAM designed the cognitive stimulation activity, titled *Bingo-Musical*, suited to the needs of individuals with ID. In this *Bingo-Musical* activity, a random song from a predefined list is first played and then a user is selected at random and is asked to identify the name of the singer for the played song. If the selected user is not able to answer, another user is chosen randomly. On the correct identification of the singer by the user, the corresponding video song is played with an applause. The caregivers selected songs that were popular in the region and hence likely to be known by the users. Two versions of the activity were designed, one without the robot and another with the robot. The same caregiver was present in both the types of the sessions.

Ethical, legal and social issues concerning the trials were identified by the *Institutional Advisory Board (IAB)* of the respective care facilities and accordingly the trials were designed and executed. A written consent to take part in the study was obtained prior to the experiments, by the residential care facilities and by the guardians of the users involved in the study.

5.1.1 Participants

The proposed cognitive stimulation sessions were performed by five different groups of users, each one led by a different caregiver. Two caregivers from FAM, two caregivers from L'Espiga Foundation and one caregiver from ASPACE Confederation participated in the study. All selected five caregivers were female. The chosen caregivers aged between 26 and 49 years (M = 38.6; SD = 9.24). The experience of the caregivers ranged between 2 and 12 years (M = 8.2; SD = 4.27). Only the two caregivers from FAM had prior work experience with the robot.

[1] Fundació Ave Maria, http://www.avemariafundacio.org/inici.html
[2] Fundació l'Espiga, http://www.fundaciolespiga.com/
[3] Confederación ASPACE, http://www.aspace.org/

The criteria for caregiver's selection were as follows:

1. Availability to participate in both the sessions (with and without the robot) at the same time of the day.
2. Familiarity with individuals of the group for at least two years.
3. Users were comfortable with the caregiver.
4. Regularly exercising cognitive stimulation activities with the users.

Each group was composed of six individuals. Twelve individuals from FAM, twelve individuals from L'Espiga Foundation and another six individuals from ASPACE Confederation participated in the study. Hence, the chosen sample consisted of 30 users (18 Females and 12 Males) aged between 26 and 69 years (*M* = 45.24; *SD* = 11.28). Users had a diagnosis of either mild (2), moderate (21) or severe (5) ID; the degree of ID in the remaining users (2) was unreported. Causes of ID were heterogeneous and in many cases unknown.

User's inclusion criteria were defined as follows:

1. Age equal to or above 18 years.
2. An official diagnosis of ID as done by *Assessment and Guidance Services for People with Disabilities (CAD Badal)* organization[4].
3. Living in the residence facilities for at least 3 years.
4. Ability to identify the singer and the song's association.
5. Their guardians have provided written consent to take part in the study.

All selected 12 users from *FAM* were familiar with the robot while, other 18 users never had any interaction with a robot.

5.1.2 Set Up

Figure 5.1 shows the experimental setup, consisting of an intervention table, along with the required multimedia devices for the activity. The users sat on either side of the table, facing each other. The caregiver was standing and moving around while monitoring and proving assistance to the users. Multimedia devices required for all the sessions were a computer system (processor, monitor, keyboard, mouse) for showing the images and videos and a

[4] Generalitat de Catalunya, http://web.gencat.cat/ca/inici/

Fig. 5.1 Experimental set up for the sessions with the robot

speaker system for playing the songs. Sessions with the robot involved a humanoid robot and a Radio-frequency identification (RFID) reader to get the input from the users and/or the caregiver during the session. All users were given the images of all the singers in the specific list of that session. The images of the singers were tagged with the high frequency (HF) RFID tags. The robot used for this study was NAO (NextGen Model H25, V4)[5] and the RFID reader was TWN3 Multi ISO transponder Reader/Writer[6]. The RFID reader was placed on a toy truck to make it appealing to the users. A high-quality audiovisual recording was obtained by placing the camera facing the robot 3 feet away from the table.

5.1.3 Procedure

The activity was explained to all the five caregivers before starting the actual sessions, while a demo session was performed with the robot. The differences in the execution of the sessions without the robot were also explained to all the five caregivers. During the actual sessions, the users were brought to the trial room by the caregiver and were given a seat at the table. The robot (in sessions with the robot) and the multimedia devices were prepared by the researcher, before the arrival of the users. After users arrived, the caregiver distributed them the RFID tagged sets of singer images. The image set exclusively consisted of the images of the six

[5] NAO, https://www.ald.softbankrobotics.com/en/cool-robots/nao
[6] Elatec RFID Systems, https://www.elatec-rfid.com/

singers associated with the songs of the session. The RFID image sets were of six different colors and were associated with a particular individual by the researcher. Such association helped the robot to identify a particular individual and hence the distribution of the tags were done under the instruction of the researcher. Then the camera recording was started and the researcher left the room. Depending upon the type of the session, the robot or the caregiver started the execution of the activity by explaining the activity to the users. The robot was programmed to autonomously execute the session with the caregiver. The execution flow of both the versions (caregiver only and caregiver + robot) and their comparative differences are given in table 5.1.

During the caregiver only version of the activity, the interaction between the user and the caregiver was natural and hence the caregivers were free to interact with the users as they desired. The duration of the trials varied between 10-20 min depending upon the response of the users. Three groups started their sessions with the robot version of the activity, while other two with the caregiver only version.

5.1.4 Evaluation

Motivated by the evaluation metrics presented in section 3.3.5, the impact of the SAR on the caregivers was evaluated on the following three aspects:

1. **Subjective Workload**

   The NASA Task Load Index (NASA-TLX) questionnaire (Hart and Staveland, 1988) was used to measure the caregiver's subjective perception of workload in the tasks with and without the robot. NASA-TLX is a multidimensional scale which provides an estimate of task's overall subjective workload, calculated as the weighted sum of six factors, including task mental, physical, and temporal demands, as well as perceptions of effort, performance, and frustration. It has been proven as an effective tool for measuring workload in several contexts and is currently one of the most widespread methods for analyzing subjective task workload (Hart, 2006).

   A Spanish version of the NASA-TLX questionnaire, implemented in an in-house Python application, was administered to the caregivers after each session. The questionnaire consisted of two parts; in the first one, all the possible pairs of factors were presented, and the participant was asked which factor in each pair contributed more to the overall workload of the task. The weight for each factor (ranging from 0 to 5) for each participant in each task was calculated as the number of pairs in which the participant selected that factor as the more determinant. In the second part of the scale, the participant was asked to rate each of the six factors in a Likert-type scale ranging

Table 5.1 Bingo Musical Activity: Caregiver only and the Caregiver + Robot Version

Caregiver only Version	**Caregiver + Robot Version**
The caregiver explains the activity to the users.	The robot explains the activity to the users.
The caregiver confirms if the participants understood the activity	The robot confirms if the participants understood the activity by listening to the yes (Spanish - *Si*) sound from the users.
The caregiver plays the audio song on the speaker via laptop and motivates the user to sing along while dancing.	The robot plays the audio song on the speaker and encourages the participants to sing along and it imitates *guitar playing*.
At the end of the song, the caregiver chooses one participant at random to identify the singer and claps for him/her. The caregiver then asks the selected participant to identify the singer by pointing the singer's image.	At the end of the song, the robot chooses one participant at random to identify the singer. The robot presents his/her picture on the monitor and claps for him/her. The robot then asks the selected participant to identify the singer by placing the image on the *truck*.
Upon correct identification, the caregiver gives an applause to the user and plays the video of the song on the monitor. The caregiver sings/dances while the video song is getting played and it encourages all the participants to sing/dance along.	Upon correct identification, the robot gives an applause to the user and plays the video of the song on the monitor. The robot sings while the video song is getting played.
In case of a wrong response, the caregiver encourages the user to try again. If the caregiver decides that the user cannot provide the response, she then selects another user.	In case of a wrong response, the robot encourages the user to try again. If the caregiver decides that the user cannot provide the response, she provides a *pass* tag to the *truck* to let the robot know that user can not answer. So, then the robot selects another user.
The activity continues in above fashion until all the songs are in the list are played.	The activity continues in above fashion until all the songs are in the list are played.

from 0 to 20. Each scale included a Spanish translation of the description of the factors as provided in (Hart and Staveland, 1988).

2. **Division of Caregivers Time**

   The time spent by the caregiver providing personalized attention to the users was annotated by two caregivers from *FAM* who had not participated in the experiment. An in-house Python application was implemented for conducting the annotations. The video recordings of each session were divided into ten-second segments. Each annotator watched the segments individually, and after each segment, the annotator classified the activity of the caregiver in that segment into one of four possible classes:

   - *User*: The caregiver is providing personalized attention to one or more users (e.g. helping the user to overcome his/her physical impairment for active participation)
   - *Activity*: The caregiver is carrying out executional aspects of the activity (e.g. setting the video player)
   - *Both*: The caregiver is both providing attention to the users and taking charge of executional aspects of the activity or
   - *None*: None of the above.

3. **Qualitative Interviews**

   Finally, in order to obtain insights on the subjective impressions of the caregiver with the robot, a questionnaire interview was designed and administered to the caregivers after they finished the last session. Their responses were recorded using a mobile phone and later transcribed by the researchers. The questionnaire interview included a first question about their opinion on whether the robot was useful in reducing the workload in this kind of tasks, with four possible answers: *Yes, a lot*; *Yes, some*; *No*; *No, actually it increases the workload*. The next four questions were open questions, in which the caregiver was asked to express her opinion on:

   - Main advantages of using the robot for this kind of task
   - Its main disadvantages
   - Possible improvements in the robot and
   - Other relevant comments.

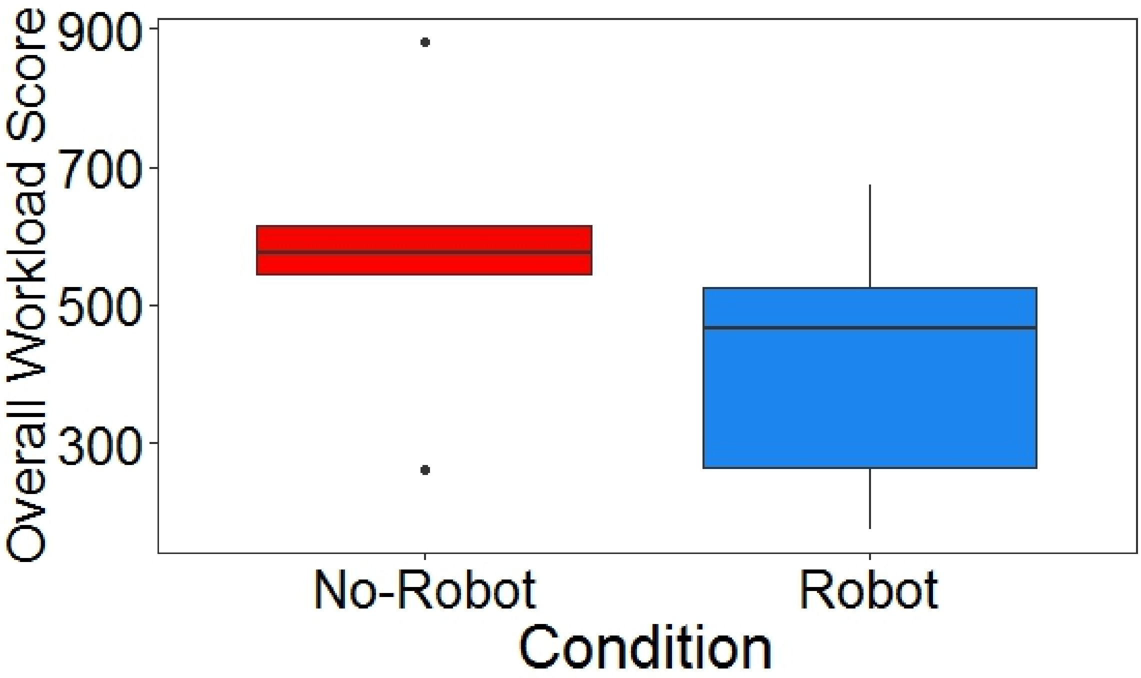


Fig. 5.2 Overall Workload

5.1.5 Results

Subjective Workload

Scores given by the caregivers to each factor in the second part of the NASA-TLX questionnaire were multiplied by 5, in order to render the 20 points original scale into a 100 point scale (Hart and Staveland, 1988). Afterwards, the score for each factor was multiplied by the weight given by the caregiver to that factor in that session, and the weighted scores of the six factors were added up in order to obtain the overall workload estimate.

The results of the NASA-TLX questionnaire suggest that the use of the robot leads to a significant reduction in the task's subjective workload (Fig. 5.2). Figure 5.2 shows the box-plot of the overall workload score, where the graph represents the minimum, maximum, median, first quartile and the third quartile in the data set and the outliers are plotted individually using the (·) symbol. In all the five cases the overall workload estimate was lower for the task with the robot than for the task without the robot. The reduction of the subjective workload among the five caregivers ranged between the 4% and 54%, with an average reduction of 28% (SD = 18%). We conducted a paired t-test comparing the overall workload score of both the versions of the activity. The result suggests that there is strong evidence of a mean decrease in overall workload score between the without the robot and the with robot version of the activity, $t(4)=3.0922$; $p<0.005$.

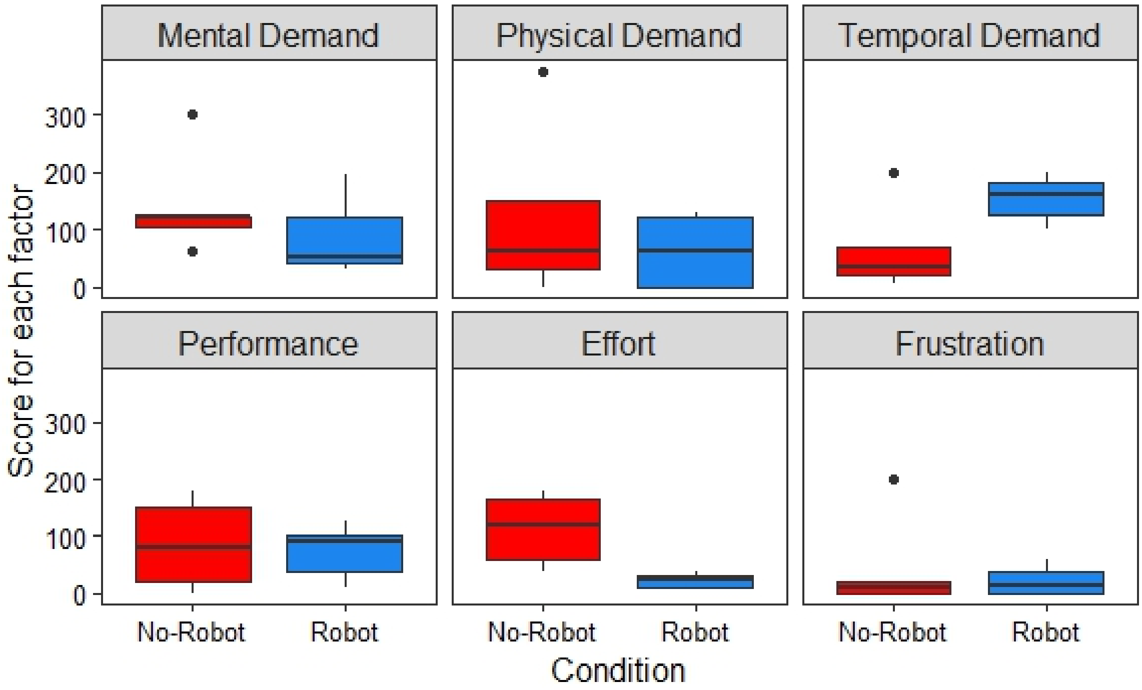


Fig. 5.3 Workload by factors

The details of the scores for each of the six factors analyzed in the NASA-TLX questionnaire (Fig. 5.3) reveal that the use of the robot impacts mostly on the mental demand of the task and on the effort required of the caregiver, while actually the use of the robot increases the temporal demands of the tasks. Table 5.2 shows the paired t-test analysis of all the six factors individually, comparing without the robot and the with robot versions of the activity. The results suggest that differences between the without robot and with robot activity was not significant in terms of the mental demands, $t(4) = 0.90843$; $p = 0.415$, the physical demands,$t(4) = 1.2608$; $p = 0.2759$, the temporal demands, $t(4) = -2.5555$; $p = 0.415$, the performance perceptions, $t(4) = 0.51427$; $p = 0.6342$ and the frustration perception, $t(4) = 0.7754$; $p = 0.4814$. By contrast, there was a significant difference in terms of effort perception, t(4)= 3.6974; $p<.05$.

Division of the Caregivers Time

Regarding the time spent with personalized attention to users, a total of 862 segments was annotated by the two annotators. Their annotations were in agreement only in 491 cases (∼57% of the sample). In order to deal with the conflicting classification of segments between the two observers, a set of rules was defined. In the cases that the two classes for a segment were *User* and *Activity*, the segment was considered as belonging to the *Both* class. If one

Table 5.2 *t*-test Analysis: NASA-TLX Score

Factor	***t***	***p***
Mental Demand	0.90843	0.415
Physical Demand	1.2608	0.2759
Temporal Demand	-2.5555	0.06294
Performance	0.51427	0.6342
Effort	3.6974	0.02088
Frustration	0.7754	0.4814

annotator classified a segment as *User* or *Activity* and the other annotator classified it as *None*, then the segment was classified as *User* or *Activity*; whereas if instead of *None* the divergent class was *Both*, the segment was assigned to the *Both* class. Finally, if the two classes for a segment were *Both* and *None*, the segment was discarded. A total of 35 segments ($\sim$4% of the sample) was discarded, remaining a total of 827 segments. Then the amount of the time dedicated to each class relative to the whole session was calculated for each session and averaged between sessions for each condition.

The results are presented in figure 5.4, showing that, on average, about 36% of the time was devoted to personalized attention to users in both conditions. Time spent in technical aspects of the activity, as well as time spent in both user and activity, was lower in the robot condition, while, the time spent in *None* label was higher in the robot condition. We conducted a paired *t*-test comparing the time devoted to each class in both the versions of the activity and the results are presented in table 5.3. There is no significant difference in terms of the percent of the time spent on *Activity* or on *User*. However, there is a significant difference (p<.05) in terms of the time spent in *Both* and *None*, although these results need to be taken with caution given the small sample size.

Qualitative interviews

In response to whether the robot is helpful with the activity, all the five caregivers answered affirmatively: three of the five caregivers replied *Yes, a lot*, while two of them chose the option *Yes, some*. Regarding the advantages of the using the robot, four caregivers agreed that the main advantage is that it is very helpful in drawing the users' attention, making them engaged with the activity, although two of the caregivers showed concern that it could be the consequence of a novelty effect, and thus it might be reduced when the users are more

Table 5.3 t-test Analysis: Time Division

Class	t	p
Activity	2.14	0.1
User	0.05	0.97
Both	3.21	0.03
None	6.43	0.003

familiar with it. Two caregivers considered that an advantage of the robot is that, since it takes charge of the technical aspects of the tasks (e.g. playing music and videos) the caregiver can focus more on providing personalized attention to the users. One caregiver stressed that all the users, independently of their conditions, could benefit from its use, while another stated that it seems that it could be applied to a broad range of activities.

The main disadvantages mentioned by the caregivers were associated with technical details of the implementation of the activity. More specifically, four of them mentioned issues related to the rhythm of the activity, which was too slow in some moments, and the need for more flexibility in the dynamic of the activity (e.g. being able to stop the activity when something unexpected happens, or to make the robot repeat instructions when some user does not understand them). Three caregivers also highlighted that the pronunciation of instructions, and especially user's names, was not correct in some cases. Two caregivers mentioned that the robot may be perceived as *cold* by some users, that it would need to be more *human*. Another two caregivers pointed out the difficulty of programming and using the robot as a disadvantage.

Improvements suggested by caregivers were related to increasing the interaction between the robot and the users, by enabling the robot to make more personalized comments, or even physically approaching each user during his or her turn in the activity. One caregiver highlighted that more movement and variation in the robot choreographies may help to engage users more, and another caregiver mentioned that if the caregiver had more control of the robot during the activity it would be easier to adapt the rhythm of the activity to user's needs.

Other comments made by caregivers included stressing the possibilities of the robot in other types of activities (such as daily activities and reinforcement routines, e.g. during lunch time); that a more personalized robot would help the users to perceive it as *one more in the group* and so boosting more social interactions; and that, even though the robot is helpful and

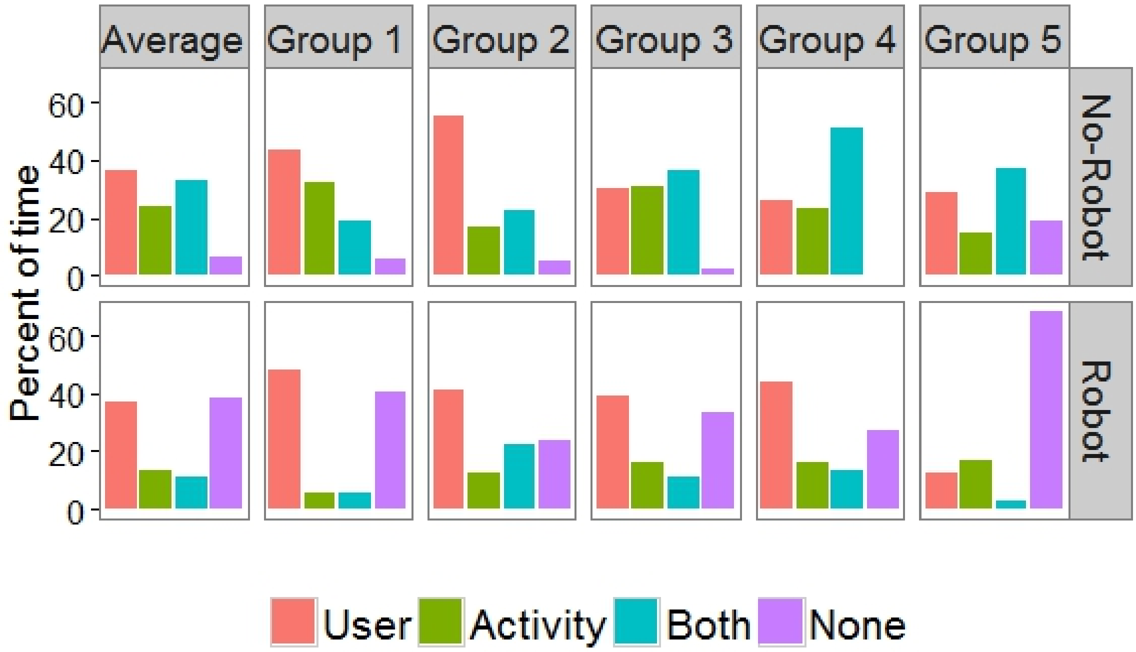


Fig. 5.4 Division of Caregivers Time

the users appreciate it, it would never be able to replace the figure of the caregiver, which performs interventions that are out of the scope of the robot.

5.2 Discussion

The present results offer a general picture of the effects of the use of robots during cognitive stimulation tasks from the point of view of caregiver's workload. The multidimensional approach of the present research has allowed exploring different dimensions of caregiver burden. As expected, the fact that the robot takes charge of the technical aspects of the tasks (e.g. playing the music and the videos) reduces task workload, operationalized either as caregiver's subjective perception, or as the proportion of time spent on the technical aspects of the task.

The results of the NASA-TLX questionnaire suggest that, even though the robot apparently increases the temporal demands of the task, there is a reduction in overall workload, consistently reported by the five caregivers, and driven by a pronounced decrease in effort and, to a lesser extent, in mental demand.

The division of caregiver's time suggested a clear reduction on time spent on technical aspects of the tasks, but no increase in the time devoted to users personalized attention.

Indeed, the more striking aspect of these results is that the time saved from technical aspects of the activity is not spent on personalized attention to users, but falls into the *None* class (time not invested in the executional aspects of the activity nor in users personalized attention). After analyzing the videos, it becomes evident that the caregivers invested this *free* time in either busy waiting for the robot to finish its part or in monitoring the group's reactions. While this deeper monitoring of users may have some advantages, a higher increase in user's personalized attention during the task would also be desirable. In this sense, a concern has already been shown that caregivers are not used to working with robots, and may need to adjust their work habits Scholtz (2002). It is likely that, as caregivers become more familiar with the use of the robot, they develop the strategies to optimize their benefits, but, in any case, institutions considering the use of robots in this kind of tasks need to consider caregivers specific training as a priority in order to maximize the benefits of such decision.

From a subjective point of view, the impressions given by the caregivers were consistently positive, not only about the effect of the robot in their workload but also about the positive effects of the robots on user's engagement with the activity. Whereas evaluating the effectiveness of the use of robots on user's performance was not the goal of this research, the views provided by the caregivers provide anecdotal evidence in the sense that the robot can positively contribute to users performance, and encourage research on this topic. The most negative aspects of the robots were related to the technical implementation of the actions done by the robot (e.g. bad pronunciation, slow rhythm) that can be easily solved in future implementations of the activity. A higher control on the pace of the activity by the caregiver may also help to reduce the increase in task temporal demands detected by the NASA-TLX questionnaire. The caregivers comments and suggestions on improvements collected through the qualitative interview may serve, thus to improve the robots functionalities according to user's needs, mainly in the sense of including more personalization, adaptability, and interactivity in the robot's behavior.

One clear limitation of the present research is the small sample size, restricted to only 5 caregivers. This was due to the characteristics of the research and the difficulty in finding institutions fitting the requirements. However, the consistency of the results (e.g. the fact that all the caregivers agree in reporting a reduction in workload, and the consistency in their comments) suggests a clear positive effect of the robot in reducing caregiver burden. Yet, it would need to be affirmed by future research with larger samples and with different types of tasks.

5.3 Conclusion

The findings of presented research provide valuable insights on reduction of burden on caregivers serving individuals struggling with a wide range of mental health concerns. However, it also has been shown that the benefits on the reduction of the caregiver's workload do not automatically translate into an increase in user's care. A possible reason for this could be the lack of familiarity and specific training with the robot among the caregivers. While the use of SAR cannot replace caregivers at residential care facilities, it has been proven that it can help them to focus their time and resources on providing better monitoring and personalized attention of the individuals.

Part II

Emotion Recognition using EDA signals

Chapter 6

MuDERI: Multimodal Database for Emotion Recognition among Intellectually Disabled Individuals

Summary. Social robots with empathetic interaction is a crucial requirement towards deliverance of an effective cognitive stimulation among individuals with Intellectual Disability (ID) and has been challenged by the absence of any particular database. In this chapter, we present a first-ever multimodal database of individuals with ID named *MuDERI*, recorded in *nearly real world* settings for analysis of human affective states. MuDERI is an annotated multimodal database of audiovisual recordings, RGB-D videos and physiological signals of 12 participants in actual settings, which were recorded as participants were elicited using personalized real-world objects and/or activities. The database is publicly available.

A crucial step towards delivering an efficient cognitive stimulation to individuals with ID by robots is to make them able to perform an emotionally adaptive behavior; i.e., to be able to detect users' feelings and to adjust the experience to fit them (Shukla et al., 2015)(Shukla et al., 2016b). However, affective state estimation among such type of users by robots has been challenged by the absence of any multimodal database for individuals with ID. There are publicly available databases for research purposes that contain naturalistic multimodal and continuous data, labeled either in terms of discrete categories or along the emotional dimensions, such as Multi-modal Affective Database for Affect Recognition and Implicit Tagging (MAHNOB-HCI) (Soleymani et al., 2012), Database for Emotion Analysis using Physiological Signals (DEAP) (Koelstra et al., 2012), SEMAINE database (McKeown et al., 2012), Belfast Induced Natural Emotion Database (Sneddon et al., 2012), etc. However, we could not use aforementioned databases for reasons mentioned below:

1. Existing databases employ *non-personalized stimulation* circumstances to stimulate emotions among subjects, while project REHABIBOTICS aims for personally significant circumstances which are *personalized stimulations*. Hence, presented database MuDERI employs *natural* setup for gathering genuine emotions among individuals.

2. Previous studies indicate that individuals with ID have Electroencephalogram (EEG) abnormalities Ünal et al. (2009) and as existing databases were recorded with healthy participants, a unique database was highly demanded by the project REHABIBOTICS.

The aim of presented work is to overcome the issues above by introducing the MuDERI database which will assist in empowering the robots with the automated emotion recognition ability among users with ID during human-robot interaction.

6.1 Method

A series of cognitive stimulation sessions among users were conducted at the *Ave Maria Foundation (FAM)*[1] for the creation of the multimodal database. *FAM* is a residential and clinical facility for users with ID. A series of meetings with the caregivers at *FAM* were done for the identification of the emotional states that the users most commonly exhibit and ease in arousal of specific discrete emotions (e.g. happiness, sadness etc.) among them, through the stimulation sessions. To the best knowledge of the authors, this is the first experiment demonstrating on-line recognition of emotional states and personalized adaptation during a cognitive stimulation activity.

Consequently, the design of tasks was favored on an approach based on the bi-factorial model of emotions. Hence, two cognitive stimulation sessions were designed for each user, one aimed to elicit positive emotions (joy), and the other aimed to elicit negative emotions (sadness or anger), from an initial neutral emotional state.

Since the aim of the experiment was to record the emotional response of the participants in *real-world* scenarios, real-world objects/activities were used to stimulate desired emotions. A list of activities and/or objects, *personally significant* for the participant and capable of provoking a desired emotion, were identified by three caregivers for each participant. These real world objects/activities were later used to provoke the desired emotions during the activity and to adapt the interaction by the robot. Only those caregivers, who have been

[1] Fundació Ave Maria, http://www.avemariafundacio.org/inici.html

taking care of these individuals for at least three years and hence, were fully aware of their behavior, participated in the identification of such objects and/or activities. Examples of tasks addressed to elicit positive emotions were: playing the user's favorite music, giving the user a candy of his/her choice, etc., whereas examples of tasks stimulating negative emotions were: discussing the user about his/her demised parent, trying to take his/her bracelet out, etc.

Ethical, legal and social issues concerning the trials were identified by the *Institutional Advisory Board (IAB)* of the *FAM* and accordingly the trials were designed and executed.

6.1.1 Participants

Participant's inclusion criteria were defined as follows:

1. Age equal to or above 18 years.
2. An official diagnosis of ID as done by *Assessment and Guidance Services for People with Disabilities (CAD Badal)* organization[2].
3. Living in the residence facilities for at least 3 years, so caregivers are familiar with them and are able to interpret their communication intent.
4. Their guardians have provided written consent to take part in the study.
5. Participants felt comfortable wearing the non-intrusive wireless physiological sensors.

The chosen sample consisted of 12 participants (10 females and 2 males) aged between 34 and 69 years (M = 49.58; SD = 10.58). Participants had a diagnosis of either moderate (4) or severe (5) ID; the degree of ID in the remaining participants (3) was unreported. Causes of ID were heterogeneous and in many cases unknown.

6.1.2 Set Up

Figure 6.1 shows the experimental setup, consisting of an intervention table which was specially designed for performing stimulation trials among the participants. The participant and the caregiver sat on either side of the table, facing each other. The intervention table is equipped with an arch above the middle of the table, where two cameras (one high-resolution video camera and one RGB-D camera) were mounted. Such a placement of the camera's allowed recording of the participant's audiovisual and RGB-D videos, without hindering the interaction between the participant and the caregiver. A Logitech C-920-C[3] camera was

[2]Generalitat de Catalunya, http://web.gencat.cat/ca/inici/
[3]http://support.logitech.com/en_us/product/c920-c-webcam

Fig. 6.1 Schematic representation of the set up

used in the audiovisual recordings and an Asus Xtion PRO LIVE[4] was used for the RGB-D recordings.

Two physiological sensors were worn by the participants and the sensors were fully wireless to cause minimal intrusion to the participant. EEG signals were collected by means of a headband sensor Emotiv Epoc[5], a wireless system that provides data from 14 EEG channels and two reference channels, with an internal sampling rate of 2048 Hz automatically filtered and downsampled to 128 Hz. Data channels collected by this device include the following positions, according to the International 10-20 system: AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, and O2. Participants' EDA signals were collected using a wireless wristband sensor Shimmer GSR+[6]. Figure 6.2 shows the devices used to obtain physiological signals. Low cost of these devices promises a wider reach. A laptop was placed behind the intervention table, hidden from the participant. It was used to receive the data from physiological sensors, video and RGB-D cameras.

(a) Emotiv Headset (b) Shimmer Device

Fig. 6.2 Physiological Sensors

[4]https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/
[5]http://emotiv.com/epoc/
[6]http://www.shimmersensing.com/shop/shimmer3-wireless-gsr-sensor

6.1.3 Procedure

During the trial, the participant was brought to the trial room by the caregiver and took a seat in front of the intervention table. The caregiver put the physiological sensors on the participant, including the Shimmer device on one wrist, and the Emotiv headset on the participant's head. Then the signal recording was started and the researcher left the room. Thus, the caregiver stayed in the room with the participant while the researchers observed and controlled the whole situation from observation area set up in the room, hidden to the participant. After that, the recording of the baseline activity was done during 30 seconds, while the caregiver talked to the participant trying to not to elicit any specific emotion, in order to provide a recording of an emotionally neutral state. The caregiver presented the object and/or activity to the participant and interacted with her in order to elicit the target emotion, while assisting the participant during all the task. The duration of the trials varied between 15-20 minutes, depending upon the setup time and activity, while actual stimulation during each trial lasted between 3 to 5 minutes.

Each individual participated in two sessions, conducted in different days, one aimed to elicit a positive emotion and the other aimed to elicit a negative emotion, and each session involved two periods: the baseline period, and the trial period. A positive reinforcement was conducted a post-session for sessions eliciting a negative emotion in the participant. This was done for a few minutes after the session until the caregiver considered that the effects of the negative emotion were not evident in the participant anymore.

6.1.4 Data Preprocessing and Annotation

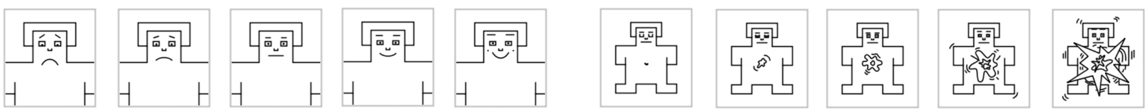
The audiovisual recordings were used as a reference signal for the synchronization of the RGB-D data and physiological signals. The EEG signals, EDA signals and kinect data were trimmed and synchronized with audiovisual recordings using the timestamps, registered earlier with the help of corresponding recording softwares. Kinect videos were trimmed using OpenNI[7]. EEG signals were trimmed and preprocessed using EEGlab software Delorme and Makeig (2004). A band-pass filter (4-45 Hz) was applied to the raw signal, and ocular artifacts were automatically removed Gomez-Herrero et al. (2006). EDA signals were trimmed using an in-house Python script and were filtered using Ledalab software Benedek and Kaernbach (2010a). The raw signals were low-pass filtered (1 Hz cutoff frequency) to remove the artifacts of the signals and smoothed with moving average method with 8 samples window. However, a lot of artifacts, mainly due to participant's movement, remain in the EDA signals,

[7]http://structure.io/openni

so researchers using the database are recommended to use advanced techniques to remove such artifacts.

The data annotation was conducted on specific moments of the audiovisual recordings. One caregiver, familiar to all the participants, selected a number of moments in each audiovisual recording in which the participant showed a neutral emotional state (no evidence of emotion) as well as moments in which he or she shows evidence of being experiencing emotions. The inclusion of neutral (not-emotional) moments in the annotation task is intended to provide annotated data from each participant that can be used as a baseline for the analysis of emotional moments. One or two neutral moments and three to six emotional moments were selected for each participant in each period for posterior annotation. The duration of all moments was about 10 seconds. Five caregivers (3 of those were involved with experiments and hence were familiar with trial intents while other 2 were not) carried out the annotation task. It was done to avoid any prejudice biasing that could affect the evaluation. The audiovisual recordings of the preselected neutral and emotional moments were presented to the caregiver using an in-house Python application. After watching each sequence, the caregivers annotated it in terms of emotional valence and arousal using a 9-points SAM scale (Morris, 1995). They also annotated each sequence in terms of the joy, sadness, and anger that in their opinion was manifest in the participant, using a 9-points Likert-type scale (Bowling, 2014).

The SAM scale is composed of two scales: valence and arousal scales (figure 6.3[8]). In the valence scale, the caregiver was required to rate the hedonic valence of the user during the task, that is, how positive or negative the user felt during the task. The left end of the scale represents the most negative experience (unhappy, sad, annoyed, unsatisfied, melancholic, angry, etc.) and the right end represents the most positive experience (happy, pleased, satisfied, etc.). In the arousal scale, the caregiver rated how excited the user has been during the task. The left end of the scale represents the calmest experience (calmed, relaxed, drowsy) and the right end represents the most arousing experience (excited, alert, anxious).



(a) Valence Scale (b) Arousal Scale

Fig. 6.3 SAM Mankins

[8]Irtel, H. (2007). PXLab: The Psychological Experiments Laboratory [online]. Version 2.1.11. Mannheim (Germany): University of Mannheim.

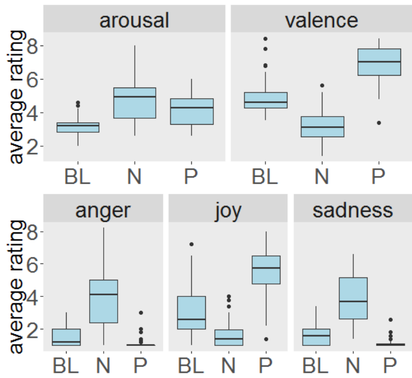


Fig. 6.4 Distribution of the average scores for each annotated variable, BL:baseline trial, N: Negative trial, P: Positive trial

6.2 Database

Table 6.1 MuDERI Database, *1 recording missed due to technical problems

	Positive Trial	Negative Trial
Electrodermal activity (EDA)	11*	11*
Electroencephalogram (EEG)	12	12
High-resolution audiovisual video	12	12
RGB-D video	12	12
Total Number of Annotated Keypoints	41	30

Table 6.1 shows the final composition of the MuDERI database. In order to check the effectiveness of the sessions for eliciting the targeted positive and negative emotions, we conducted a series of one-way ANOVAs, one for each rated variable (valence, arousal, joy, sadness, anger), taking as an independent variable the type of period (baseline, trial-positive, or trial-negative) and as the dependent variable the mean scores from the five caretakers in each annotated segment. Since the variance of the mean ratings was not equal through the three periods, we included the Welch correction in the ANOVA and the post-hoc pair comparisons were conducted using the Games-Howell approach. The results (table 6.2) show that in all cases there were significant differences between the average scores obtained in the three types of periods (baseline, trial-positive, or trial-negative). The only exception was the case of the arousal scores; although there was a difference between either the positive or the negative trial and the baseline, there was no significant difference between the positive and the negative trials.

Provided that two of the caretakers were not familiar with the participants, we also analyzed whether their scores were similar to the caretakers that were familiar to the users.

Table 6.2 ANOVA Analysis,BL:baseline period, N: Negative period, P: Positive period

One-way ANOVA			**Post-Hoc Comparisons**		
Annotated Variable	**F**	***p***	**Correlation**	***t***	***p***
			BL-N	5.93	<.001
Arousal	26.26	<.001	BL- P	5.96	<.001
			N- P	1.98	.13
			BL-N	6.51	<.001
Valence	98.72	<.001	BL- P	8.56	<.001
			N- P	12.71	<.001
			BL-N	6.9	<.001
Anger	67.17	<.001	BL- P	2.86	.015
			N-P	7.81	<.001
			BL-N	4.36	<.001
Joy	75.37	<.001	BL- P	8.13	<.001
			N-P	12.78	<.001
			BL-N	7.56	<.001
Sadness	88.32	<.001	BL- P	5.09	<.001
			N-P	9.75	<.001

We conducted a series of *t*-test comparing the average scores provided by the familiar caretakers in each variable to the average scores provided by the caretakers non-familiar with the users. The results suggest that, in general, the caretaker unfamiliar to the users provided more positive scores, as shown by a higher average valence, $t(113)=6.34$; $p<0.001$, a higher average joy, $t(113)=3.88$; $p<0.001$, and a lower sadness, $t(113)=5.37$;$p<0.001$. By contrast, they provided lower average ratings of arousal, $t(113)=9.68$; $p<.001$, and there was no significant difference in the anger ratings, $t(113)=0.09$;$p=0.92$.

The means of the scores for each variable are (Fig. 6.4), demonstrating the effectiveness of the sessions in eliciting the desired emotional states.

6.3 Conclusion

The results show that the experiment achieved the goal of eliciting a range of emotions on participants. Hence, the present research was successful in creating a multimodal database that can fuel research on *Emotional Adaptive Behavior* of robots for the cognitive stimulation of users with ID, by filling a gap in the availability of *nearly real-world* data from individuals with ID. Researchers in the field may benefit from using the database, which is publicly available by request[9].

[9]https://institutorobotica.org/en/investigation/muderi-dataset/

Chapter 7

Efficient Wavelet-Based Artifact Removal for Electrodermal Activity in Real-World Applications

Summary. Online monitoring of electrodermal activity (EDA) may serve as an economical and explicit source of information about actual emotional state and engagement level of users during their interaction with information and communications technologies (ICT) applications in *real-world* situations. In such contexts, however, EDA signal is affected by motion artifacts that introduce noise in the signal and can make it unusable. As the scope of movement minimization during EDA data acquisition is limited, this scenario demands online methods for detection and correction of artifacts with low computational cost. We propose an efficient wavelet-based method for artifacts attenuation while minimizing distortions, using a stationary wavelet transform (SWT) modeling the wavelet coefficients as a Laplace distribution. The proposed method was tested on EDA recordings from publicly available driver dataset collected during real-world driving, and containing a high number of motion artifacts, on EDA recordings from the MuDERI dataset and the results were compared to those of three state-of-the-art methods for EDA signal filtering. In addition, the proposed method was tested for the online filtering of EDA signals collected while twelve volunteers conducted tasks designed to elicit various stress states. The results evidenced that the prediction of arousal states can be significantly improved after motion artifacts removal and that the proposed method outperforms existing approaches and it has a lower computational cost. Taken together, these results evidence the effectiveness of the proposed method for online EDA filtering in real-world scenarios.

The availability of wearable devices able to measure online EDA provide the opportunity for real-world data collection in a comfortable manner in real-life scenarios outside the lab. This allows broadening the scope of EDA analysis by simply gathering information on user's psychological state to the online adaptation of the system according to the user's cognitive and emotional states. The advantages of such online adaptation are especially remarkable in the cases of the users with acute problems to identify, report, and regulate their emotions, such as children or individuals with intellectual developmental disabilities (IDD). In this sense, novel cognitive stimulation systems based on the robots or virtual reality may greatly benefit from an online input of the user emotional state from EDA signal (e.g. (Liu et al., 2008)).

However, compared to laboratory studies, the online collection and analysis of EDA signals in real-life context involves new challenges related to signal processing. While in lab experiments the participants are usually asked to not to move the hand to which the electrodes are attached, hand's movement can be hardly controlled in real-life settings. As a consequence of such movement, a partial detachment of the electrodes or pressure over them may occur, leading to the appearance of artifacts in the signal (Boucsein, 2012). If these artifacts remain in the signal when it is analyzed they can easily be misinterpreted and skew the analysis (Taylor et al., 2015); for example, they may be mistaken for an skin conductance response (SCR) indicating an increased stress, especially when the analysis is conducted automatically. Even in a context in which not much movement should be expected, the presence of such motion artifacts can be critical in the case of users whose ability to control the level of movement is limited (e.g. individuals with IDD).

Methods used in previous research to correct artifacts mainly consist of exponential smoothing (Hernández et al., 2011) and low-pass filtering (Ming-Zher Poh et al., 2010). Over the last two decades, wavelets have already proven their significant value in signal processing and image. Wavelets have also been found suitable for EDA activity modeling, given its non-stationary behavior (Lima et al., 2010). Wavelet-based sophisticated de-noising of motion artifacts have been widely used in research (Chen et al., 2015; Junli et al., 2007; Molavi and Dumont, 2012; Swangnetr and Kaber, 2013) and has offered better results due to the good localization property of the wavelet transforms (Molavi and Dumont, 2012). On the other hand, other sophisticated methods, not based in wavelets, have also been proposed. Recently, Greco et al. (2016) proposed a deconvolution based approach (called cvxEDA) using Maximum a Posteriori (MAP) estimation and convex optimization which provided a decomposition of the EDA that is robust to noise.

However, the aforementioned techniques present several important limitations when used for online filtering of the EDA signals, such as the following:

1. Exponential smoothing (Hernández et al., 2011) and low pass filtering based denoising (Ming-Zher Poh et al., 2010) methods are not able to atone the unexpectedly occurring artifacts which have higher values than EDA and undiscriminated filtering of the whole signal also distorts the EDA signals without artifacts (Chen et al., 2015).

2. Denoising with the traditional DWT wavelet transform (Swangnetr and Kaber, 2013) can exhibit visual artifacts due to lack of translation invariance and "pseudo-Gibbs" oscillations are especially pronounced in the vicinity of discontinuities (Coifman and Donoho, 1995).

3. Estimation of the noise level σ as proposed by Swangnetr and Kaber (2013) is based on the data collected during the rest period. This type of noise level estimation is an off-line and static measure and is an overhead cost on denoising. Moreover, as the actual nature of noise in the noisy signal is dynamic, hence the noise level estimation needs to be done online.

4. Gaussian mixture based modeling for the distribution of the wavelet coefficients (Chen et al., 2015) requires estimation of three model parameters, γ_j the mixture parameter, σ_j^2 and $c^2\sigma_j^2$ the variances of the two Gaussians (equation 7.1) which demand employment of iterative algorithms such as Expectation Maximization (EM) algorithm. These algorithms have a high computational complexity of $O(n+2n^2)$ for 2 mixture of Gaussians of one-dimensional data as in the current case, where n is the number of data samples (Chen et al., 2007).

$$\tilde{d}_j \sim \gamma_j N(0,\sigma_j^2) + (1-\gamma_j) N(0, c^2\sigma_j^2) \tag{7.1}$$

5. The cvxEDA method (equation 7.2) depends on convex Quadratic Programming (QP) (Greco et al., 2016) which demands polynomial time algorithms, such as interior point algorithms, for a solution. The time complexity of convex QP solving algorithms is $O(n^3)$ (Boyd and Vandenberghe, 2004), where n is the number of data samples. Denoising in online scenarios can be limited by the computational complexities of Chen et al. (2015) and Greco et al. (2016) methods.

$$y = Mq + Bl + Cd + \varepsilon \tag{7.2}$$

Whereas in the offline analysis of EDA signals visual inspection allows identifying and removing the parts of the recording containing artifacts, an online adaptation of systems based on EDA signal requires automated methods for such purpose. These methods not only need to be accurate enough to provide a good signal quality but also need to be computationally affordable enough to work online.

The aim of the proposed work is to present a wavelet-based method for filtering motion artifacts in EDA signal that fits these requirements. Experimental results described in section 7.2, 7.3.3 and 7.5.3, demonstrate the benefits of the algorithm through a comprehensive comparison with Swangnetr and Kaber (2013), Chen et al. (2015) and Greco et al. (2016) methods.

7.1 Method

De-noising with the traditional wavelet transforms can exhibit visual artifacts due to lack of translation invariance (Coifman and Donoho, 1995). Stationary Wavelet Transform (SWT) is redundant, linear and hence shift invariant in comparison to the Discrete Wavelet Transform (DWT) (Nason and Silverman, 1995). SWT also provides better sampling rates in the low-frequency bands compared with a standard DWT (Nason and Silverman, 1995).

The literature suggests that the selection of the mother wavelets can be done based upon their resemblance to either the shape of the signal (Gupta et al., 2005) or the shape of the typical motion artifact (Krishnaveni et al., 2006). Since EDA signals have asymmetric nature, asymmetrically shaped Daubechies (dbN) wavelets have been used to analyze them (Laparra-Hernández et al., 2009). Swangnetr and Kaber (2013) suggest that $db3$ is the most appropriate choice of the mother wavelet to represent the EDA signal. $Haar$ wavelet has also been used for detecting edges and sharp changes, commonly seen in motion artifacts (Chen et al., 2015). In addition, Chen et al. (2015) suggest that $Coiflet3$ wavelet can have potential as the basis function since it resembles the shape of the typical motion artifact. Hence, to analyze the effects of the different basis functions, we examine the $db3$, $haar$, and the $coiflet3$ wavelets separately, as a mother wavelet during EDA signal denoising.

Since EDA artifacts may result from the recording procedure or from the physiological responses, hence, artifact removal in EDA for above types is discussed separately:

7.1.1 Artifacts Originating from Recording

A main source of artifacts is the power line noise resulting from the AC frequency input of 50 Hz (in Europe). Since the typical frequency of the EDA signal is 0.0167-0.25 Hz (Dawson et al., 2007), this type of artifact can be treated as high-frequency noise and hence can be removed as low-pass filtering.

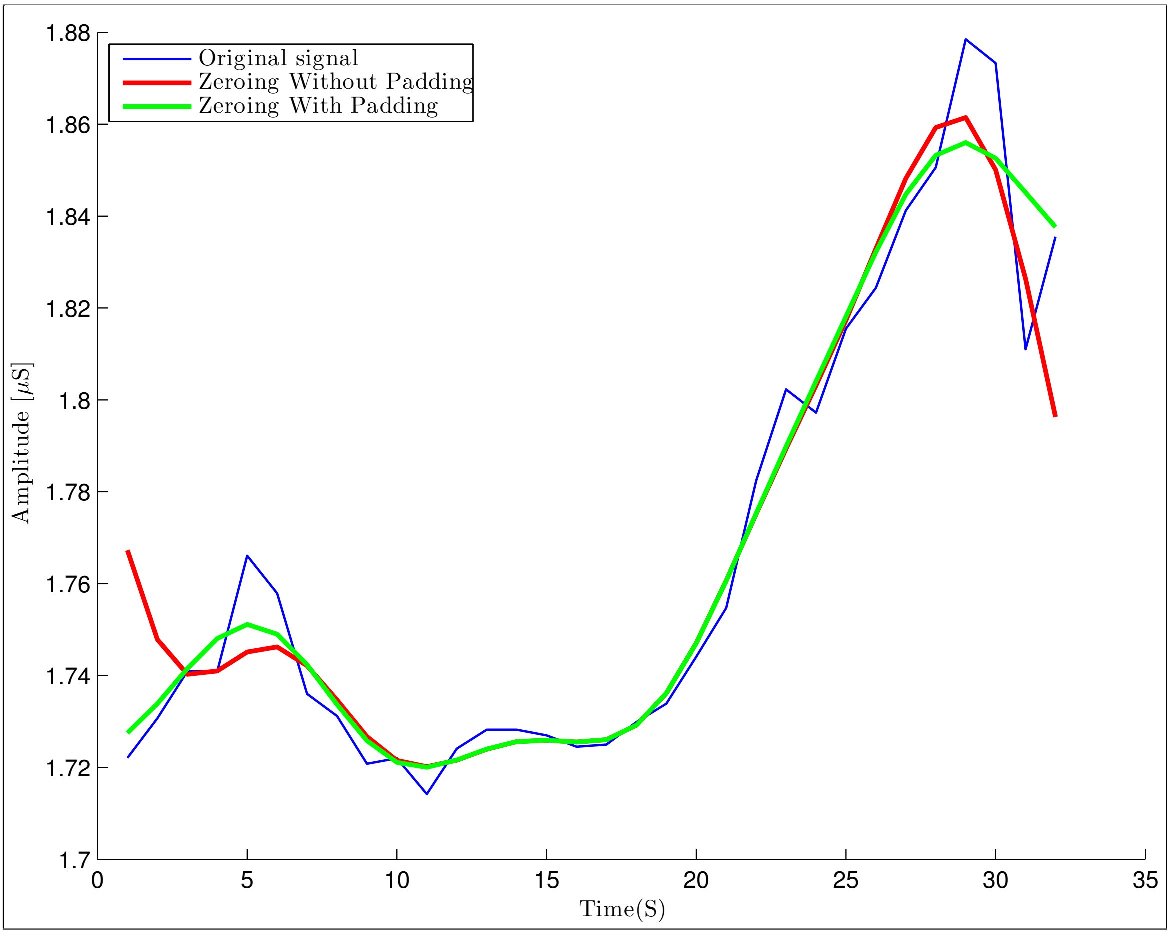


Fig. 7.1 Low pass filtering using SWT

In the proposed approach, EDA signal is modeled via SWT and the level of decomposition J is set to $(log_2 freq + 2)$, where $freq$ is the signal frequency. Detail coefficients representing high frequency ($>$ 0.5 Hz) of the signal (i.e. noise) are set to zero, thereby low pass filtering the EDA signal (Nason and Silverman, 1995). Inverse SWT is then applied to the coefficients to recover the signal.

The low pass filtering via this method distorts the signal at the beginning and at the end (figure 7.1). To get rid of the distortion, a padding can be applied at the beginning and at the end of the signal before low pass filtering. After applying the extension, the low pass filtering preserves the signal.

7.1.2 Physiological Response Based Artifacts

The most important source of physiologically based artifacts are motions, which includes skin movements beneath the electrodes and muscular activity occurring in non-recording locations (Boucsein, 2012). Swangnetr and Kaber (2013) addressed the filtering of such artifacts by using a zero-mean Laplace distribution modeling (equation 7.3).

$$\tilde{d}_j \sim Laplace(0, b) \tag{7.3}$$

The Laplace probability density function can be represented by equation 7.4, where μ is location parameter and b is a scale parameter. An estimator of μ is the sample median $\hat{\mu}$, which is equal to zero for the distribution of wavelet coefficients. Hence, the wavelet detail coefficients can be modeled using zero-mean Laplace distribution as indicated by equation 7.3.

$$f(x) = \frac{1}{2b} exp\left(-\frac{|x-\mu|}{b}\right) \tag{7.4}$$

Chen et al. (2015) used a Gaussian mixture based modeling of wavelet coefficients for removing such artifacts (equation 7.1). As is evident from equation 7.1, wavelet coefficient model proposed by Chen et al. (2015) requires estimation of three model parameters, γ_j the mixture parameter, σ_j^2 and $c^2\sigma_j^2$ the variances of the two Gaussians, whereas as suggested by equation 7.3 Laplace distribution modeling requires estimation of a single scale parameter b. Hence, computational complexity of using the Laplace distribution is significantly lower than the Gaussian Mixture Model (GMM) distribution. Figure 7.2 displays a plot of the empirical cumulative distribution function (cdf) for the wavelet detail coefficients and the corresponding theoretical cdf for the Laplace and GMM distribution along with the lower and upper confidence bounds for 5% confidence interval.

As can be seen in figure 7.2, both Laplace and GMM distributions model well the wavelet coefficients within the 5% confidence intervals. Wavelet detail coefficient obtained from a sample baseline EDA signal of a publicly available dataset (Shukla et al., 2016a) was tested for fitting against Laplace and GMM distribution using the Kolmogorov-Smirnov test. Results confirmed that the detail coefficient distributions were not significantly different from neither the Laplace distribution nor the GMM distribution ($p > 0.05$). These results reconfirm the propositions presented by Swangnetr and Kaber (2013) about Laplace distribution fit and the propositions presented by Chen et al. (2015) about the GMM distribution fit for modeling of wavelet detail coefficients.

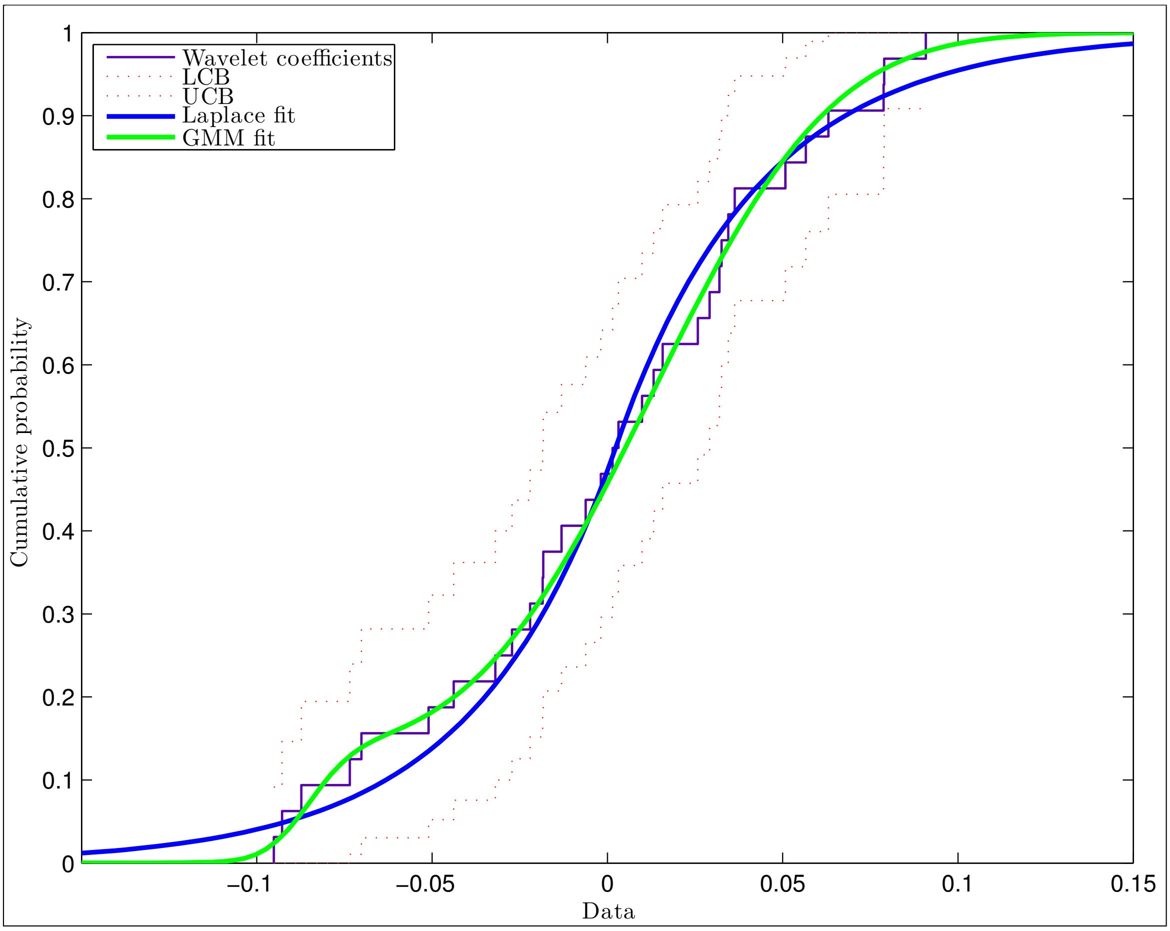


Fig. 7.2 CDF of wavelet coefficients with Laplace and GMM distribution

Laplace fitting of wavelet coefficient requires an estimation of a single scale parameter b, which can be estimated for each level j from the wavelet coefficients of the original signal d, using Maximum-likelihood estimation as per equation 7.5 (Sahli et al., 1997).

$$\begin{aligned}\hat{b} &= \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{\mu}| \\ &= \frac{1}{N}\sum_{i=1}^{N}|x_i|, (\because \hat{\mu} = 0)\end{aligned} \tag{7.5}$$

As $\hat{\mu}$ of wavelet distribution is zero, equation 7.5 reduces to later form.

Let δ be the proportion of motion artifacts in original signal d, then the thresholds of wavelet shrinkage to remove the noise coefficients from the signal can be obtained by equation 7.6.

$$\Phi(T_{low}) = 1 - \Phi(T_{high}) = \delta/2 \tag{7.6}$$

where Φ is the cumulative distribution function of the Laplace distribution, and T_{low} and T_{high} are the thresholds. It can be shown that for a zero mean Laplace distribution, T_{low} and T_{high} can be obtained using equations 7.7 and 7.8.

$$T_{low} = -T_{high} \tag{7.7}$$

$$T_{high} = \hat{b} * log_e(\delta) \tag{7.8}$$

where $\hat{b}$ is the estimated scale parameter of Laplace distribution.

Based on the points above, a SWT based denosing with $db3/haar/coiflet3$ as the mother wavelet and using Laplace distribution for wavelet coefficients modeling is employed for removing motion artifacts from EDA. The procedure is described below:

1. The highest level wavelet detail noise coefficients of the low pass filtered EDA are modeled with the Laplace distribution and the scale parameter $\hat{b}$ of Laplace distribution is estimated using equation 7.5. The T_{low} and T_{high} thresholds of wavelet shrinkage are calculated using equations 7.7 and 7.8.

2. Motion artifacts are removed from the EDA signal as per equation 7.9.

$$\hat{d_j^k} = \begin{cases} \hat{d_j^k} & T_{low} \leq \hat{d_j^k} \leq T_{high} \\ 0 & otherwise \end{cases} \tag{7.9}$$

3. Steps 1 and 2 are repeated on the wavelet coefficients of all higher levels and then inverse SWT wavelet transform is applied to the thresholded wavelet coefficients to obtain the fully denoised EDA signal.

7.2 Computational Complexity

To assess the computational complexity of the proposed algorithm, it was compared with the computational complexity of Chen et al. (2015) and Greco et al. (2016) methods. A Matlab implementation was performed for the proposed method and Chen et al. (2015) method, while Matlab implementation of the author's version was used[1] for Greco et al. (2016) method. Repeated measurements for 10,000 times were performed for all the methods while processing a sample EDA signal from a publicly available dataset (Shukla et al., 2016a).

[1]https://github.com/lciti/cvxEDA

Table 7.1 Computational Complexity for Evaluated Methods

Method	**Big O**	**Running Time (mS)**	**Parameters Required**
Chen et al. (2015)	$O(n+2n^2)$	14.931	3
Greco et al. (2016)	$O(n^3)$	21.685	3
Proposed Method	$O(n)$	**0.017**	**1**

The Matlab implementations of all the three methods were executed on an Intel®Core™ i7-4790 CPU@3.60GHz×8 processor with 8 GB of RAM running 64-bit Ubuntu 14.04 LTS operating system. Since the proposed method uses the same Laplace distribution for modeling of wavelet coefficients, as demonstrated in Swangnetr and Kaber (2013), computational complexities for both these methods are the same and hence are not compared.

Table 7.1 presents the theoretical computational complexity in terms of *Big O*, the running time of the compared algorithms in milliseconds and the number of parameters estimation required for each method. As can be seen from equation 7.5, the running time performance of the Laplace fitting method will grow linearly and in direct proportion to the size of the input data set, hence the computational complexity of Laplace fitting method is $O(n)$. On the other hand, parameters estimation for GMM distribution fitting of wavelet distribution coefficients employs the EM algorithm for fitting a 2 mixture of Gaussians which has the computational complexity of $O(n+2n^2)$(Chen et al., 2007). Similarly, cvxEDA method proposed by Greco et al. (2016) depends on convex Quadratic Programming (QP) which demands polynomial time algorithms, such as interior point algorithm, for a solution and has the time complexity of $O(n^3)$ (Boyd and Vandenberghe, 2004). Since $O(n) < O(n+2n^2) < O(n^3)$, the proposed method has the lowest computational complexity followed by Chen et al. (2015) method and Greco et al. (2016) method.

The reported running time is the median of the 10,000 repeated measurements. It can be seen that method by Chen et al. (2015) requires estimation of three model parameters, γ_j the mixture parameter, σ_j^2 and $c^2\sigma_j^2$ the variances of the two Gaussians (equation 7.1). cvxEDA approach by Greco et al. (2016) also requires estimation of optimal values for three parameters, q an auxiliary variable to find sudomotor nerve activity (SMNA), l the vector of spline coefficients, d a 2×1 vector with the offset and slope coefficients (equation 7.2). However, the proposed method requires estimation of only a single scale parameter b. Moreover, the computational cost of the proposed method is the lowest among all the three methods. In terms of execution time, it is around 863 times faster than Chen et al. (2015) method and is around 1254 times faster than Greco et al. (2016) method.

7.3 Off-line Analysis I: Driver Dataset

7.3.1 Driver Dataset

In order to test the proposed method, we used a sample of EDA signals obtained from a publicly available dataset of physiological signals collected from participants while driving on routes in city streets and highways in the area of Boston, Massachusetts (Healey and Picard, 2005). The authors of this dataset recorded, together with other physiological signals, EDA data from 24 drives with duration between 50 and 90 minutes, with a sampling rate of 16 Hz. However, since the recordings from 9 of those participants were missed, the actual sample contains about 1207 minutes of EDA recordings from 15 participants. The main reason for choosing this dataset for testing our filtering algorithm, beyond its availability, was that it was recorded during real driving. Therefore, it exemplifies well the type of motion artifacts that can be found when recording EDA in real-world scenarios.

Each signal was divided into non-overlapping 8 seconds segments, resulting in a total of 9056 segments. Eight seconds segment was used since it is adequate to capture the important features of the EDA signal (Dawson et al., 2007). A researcher with experience in the analysis of EDA signals visually inspected the signals and annotated which segments contained motion artifacts. Given the smooth waveform of EDA signal and its relatively slow change in magnitude (e.g. rise of an SCR takes 1 to 3 seconds in reach a peak), as well as its restricted range of variability (usually in a range of 1 to 3 *mS* for a given subject) (Dawson et al., 2007), experts can easily detect artifacts in the signal due to motion. Indeed, visual inspection of the signal is a standard procedure in research using EDA (for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, 2012). The expert annotated as noisy those segments containing "sharp" changes in the EDA signal, which does not represent the usual behavior of EDA and are likely to be the product of electrodes motion or partial detachment. A total of 715 segments were considered as containing motion artifacts, while the remaining 8341 segments were considered as motion artifact-free.

7.3.2 Analysis Procedure

The approach for testing the quality of the offline filtering was similar to the one described in Chen et al. (2015). It involves analyzing separately the performance of the filter for noisy and not-noisy segments. For the noisy segments, a metric of the algorithm's artifact power attenuation (APA) was calculated, while for the not-noisy segments, we calculated the normalized mean square error (NMSE) (Molavi and Dumont, 2012). This way, an overall judgment of the algorithm's performance can be made based on quantitative measures of both

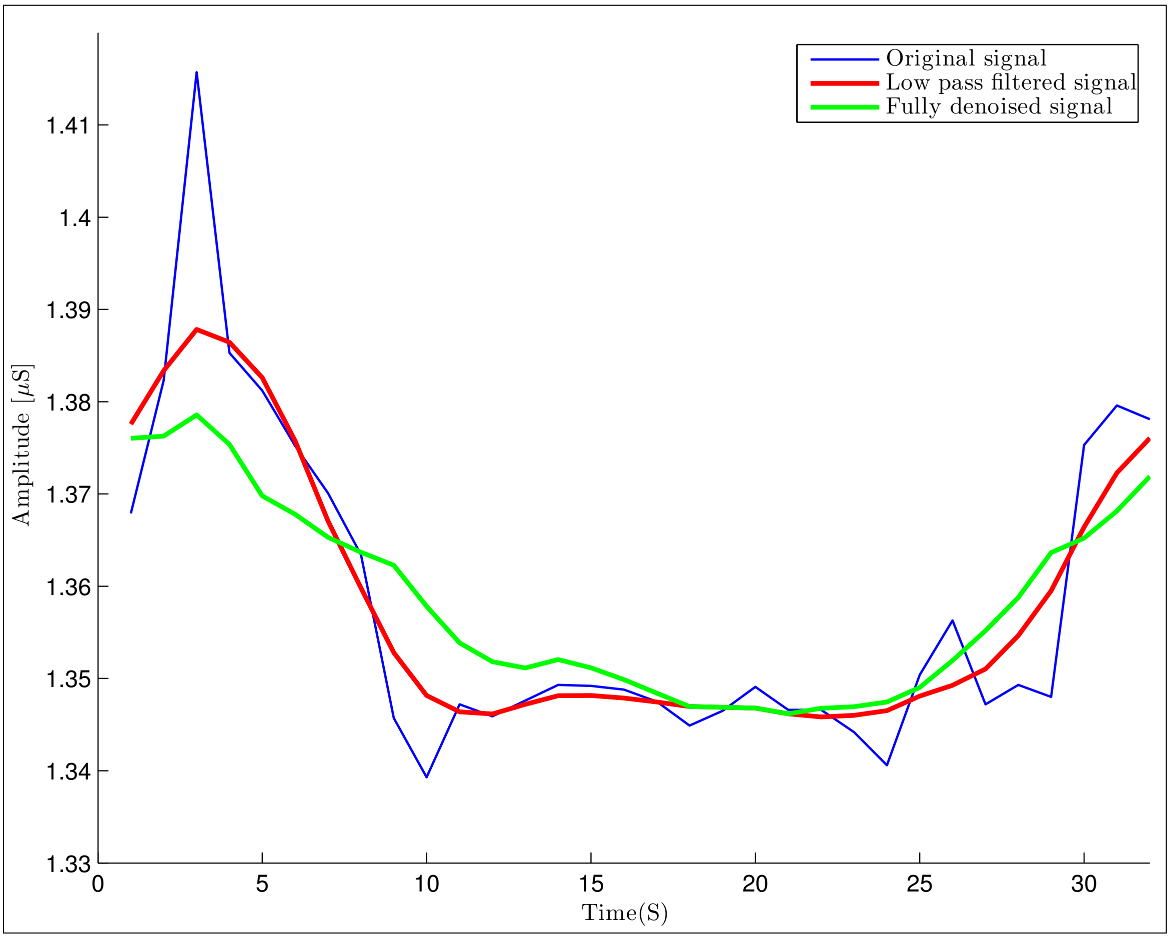


Fig. 7.3 Original EDA and denoised signals

artifact power reduction on noisy signals and distortion introduced by the filter in artifact-free signals.

We adapted the equations from Molavi and Dumont (2012) and Chen et al. (2015) to better suit the signals of short duration, by eliminating the $\sum$ over artifacts as short duration signals do not carry multiple artifacts. Therefore, APA and NMSE were defined as

$$APA = 10log_{10}\frac{Var[y]}{Var[\tilde{y}]} \tag{7.10}$$

$$NMSE = 10log_{10}\frac{[y-\tilde{y}]^2}{[y-\bar{y}]^2} \tag{7.11}$$

Where y represents the original signal, $\tilde{y}$ represents the filtered signal, and $\bar{y}$ represents the mean value of the y signal. We calculated the APA values for each of the (715) segments with artifacts and the NMSE value for each of the (8341) artifact-free segments, for the output of each of the three versions of the proposed method ($db3$, *haar*, and $coiflet3$), as well as for the same segments filtered following the approaches by Chen et al. (2015), the Swangnetr and Kaber (2013) method, and the Greco et al. (2016) approach. Concerning Swangnetr and Kaber (2013) method, the σ of noise in the raw signal and hence the threshold of wavelet shrinkage was estimated based on the 15-min rest periods at the beginning of the drive (Healey and Picard, 2005). For the cvxEDA parameters Greco et al. (2016), the default values as suggested in the code provided by the authors, which are $\tau_0 = 0.7$, $\tau_1 = 10$, $\alpha = 0.008$ and $\gamma = 0.01$ were employed for all the subjects. Concerning the Chen et al. (2015), and the proposed method, artifact proportion δ was set to 0.10 for all subjects as used by researchers in (Chen et al., 2015).

7.3.3 Results

Table 7.2 shows the median values of the APA and NMSE metrics for each of the compared methods. In terms of artifact power reduction, as expected, the median APA evidences that all the proposed methods performed better than the approach by Swangnetr and Kaber (2013). However, the APA for $db3$ and $coiflet3$ approach is lower than those of the approaches by Chen et al. (2015) and Greco et al. (2016). The *haar* version of our method outperforms Chen et al. (2015) method and is close to the performance of the Greco et al. (2016) method. In order to check the consistency of this difference, we conducted two paired samples t-test for comparing the performance of the *haar* version with the performance of the Chen et al. (2015) approach and Greco et al. (2016) approaches in terms of APA. The results of the t-tests showed that the performance of our *haar* method is significantly better than the Chen et al. (2015) approach ($t(712) = 7.38, p < .001$), while the performance of the *haar* version and the Greco et al. (2016) approach was not significantly different at the $p < .05$ level ($t(712) = 1.95, p = .05$).

Table 7.2 Median of NMSE and APA for Evaluated Methods

Method	**APA (dB)**	**NMSE (dB)**
Chen et al. (2015)	5.284	-3.085
Greco et al. (2016)	5.887	-1.885
Proposed with $db3$	4.808	-3.515
Proposed with *haar*	5.712	-3.678
Proposed with $coiflet3$	4.766	-3.665
Swangnetr and Kaber (2013)	1.624	-11.625

Regarding the NMSE values, Table 7.2 shows that all the three versions of our method have lower NMSE values, indicative of a smaller distortion introduced in the artifact-free signals, than the approaches Chen et al. (2015) and the Greco et al. (2016), being the *haar* solution the one obtaining the lowest NMSE value among them. This difference is statistically significant at the $p < .05$ level in the case of the comparison between the *haar* version and the Greco et al. (2016) method ($t(8340) = 13.67; p < .001$), but not in the comparison between the *haar* version and the Chen et al. (2015) method ($t(8340) = 1.82; p = .07$). By contrast, the approach by Swangnetr and Kaber (2013) is significantly better than our *haar* version in terms of NMSE ($t(8340) = 95.46; p < .001$).

Taken together, the present results indicate that the *haar* version of the proposed method is the one offering the highest advantages: it reduces artifact power similarly to the most powerful of the three state-of-the-art methods, while introducing much less distortion in the artifact-free segments. Compared to the approach by Chen et al. (2015), our algorithm outperforms it in terms of APA while no significant difference is found in terms of NMSE. Finally, while the approach by Swangnetr and Kaber (2013) is much more respectful of the artifact-free signals, it's potential to reduce artifact power is also much lower than our *haar* version. Thus, these results suggest that the *haar* version provides a good balance between a high artifact power reduction and a low distortion introduced in the artifact-free signals, which together with its lower computational cost, make it the optimal solution among the existing ones.

7.4 Offline Analysis II: MuDERI Dataset

We also tested the proposed method on EDA signals of the previously generated MuDERI dataset. The signal of each of the 100 segments annotated in the MuDERI dataset was downsampled to 4 Hz and the duration of the segment in 8 seconds was selected for each annotated segment from the MuDERI dataset.

In order to compare the performance of the different filters against MuDERI dataset, the proposed method was applied to filter each segment, as well as the methods proposed by (Chen et al., 2015) and (Swangnetr and Kaber, 2013). For each of the resulting filtered signals, the two metrics APA (7.10) and NMSE (7.11) were applied. A researcher experienced in the analysis of EDA signals inspected the signals and labeled each segment containing motion artifacts. A total of 67 segments were considered as containing motion artifacts, while 33 segments as artifact-free.

It was hypothesized that mother wavelet representing the signal based upon their resemblance to the signal form ($db3$) will better prevent the distortion in the denoised signal, while mother wavelets representing the signal based upon their resemblance to the motion artifact (*haar* and $coiflet3$) form will better attenuate the artifacts from the noisy signal.

7.4.1 Experimental Results

Figure 7.4 compares the denoising methods proposed by (Swangnetr and Kaber, 2013) with the currently proposed method using $db3$ as the mother wavelet signal on the same signal. It shows that the signal filtered with the currently proposed method provide better filtering in comparison to the results of the method proposed by (Swangnetr and Kaber, 2013).

In order to assess whether the performance of the proposed methods is equivalent to the performance of the (Swangnetr and Kaber, 2013) method, the APA and NMSE metrics for each method were calculated and paired t-test were applied in order to analyze if the difference between them is statistically significant. Moreover, effect size estimates (Cohen's d) were also calculated to analyze the magnitude of the difference between the outputs of the various methods in terms of the two suggested metrics (Cohen, 1988). First, the $db3$ version of proposed method was compared to the method proposed by (Swangnetr and Kaber, 2013), since both of them used the $db3$ as the mother wavelet that resembles the EDA signal and therefore are probably better preventing the distortion in the denoised signal. Table 7.3 presents the median of the two metrics applied to assess the filter performance across all the recordings.

The results of the t-tests showed that, while there is no significant difference between these two methods in terms of the APA values ($p > 0.05$), the difference in the NMSE values is significant ($t(32) = 4.31, p < 0.001$) and the effect has a medium size (Cohen's $d = 0.75$). That is, while both methods reduce artifact power in a similar way, the proposed method with $db3$ as a basis function, introduce significantly lower distortion in the artifact-free signals.

Figure 7.5 compares the denoising methods proposed by (Chen et al., 2015) with the currently proposed method using $coiflet3$ as the mother wavelet signal and with the currently proposed method using *haar* as the mother wavelet.

The next step was to compare the versions of the proposed method that are based on a mother wavelet that better fits the artifacts, so it is assumed that it should have a greater

Table 7.3 Median of NMSE and APA for evaluated methods

Method	*I*	*II*
APA (dB)	1.1775	1.412
NMSE (dB)	-17.312	-11.342

I–Proposed with $db3$ *II*–(Swangnetr and Kaber, 2013)

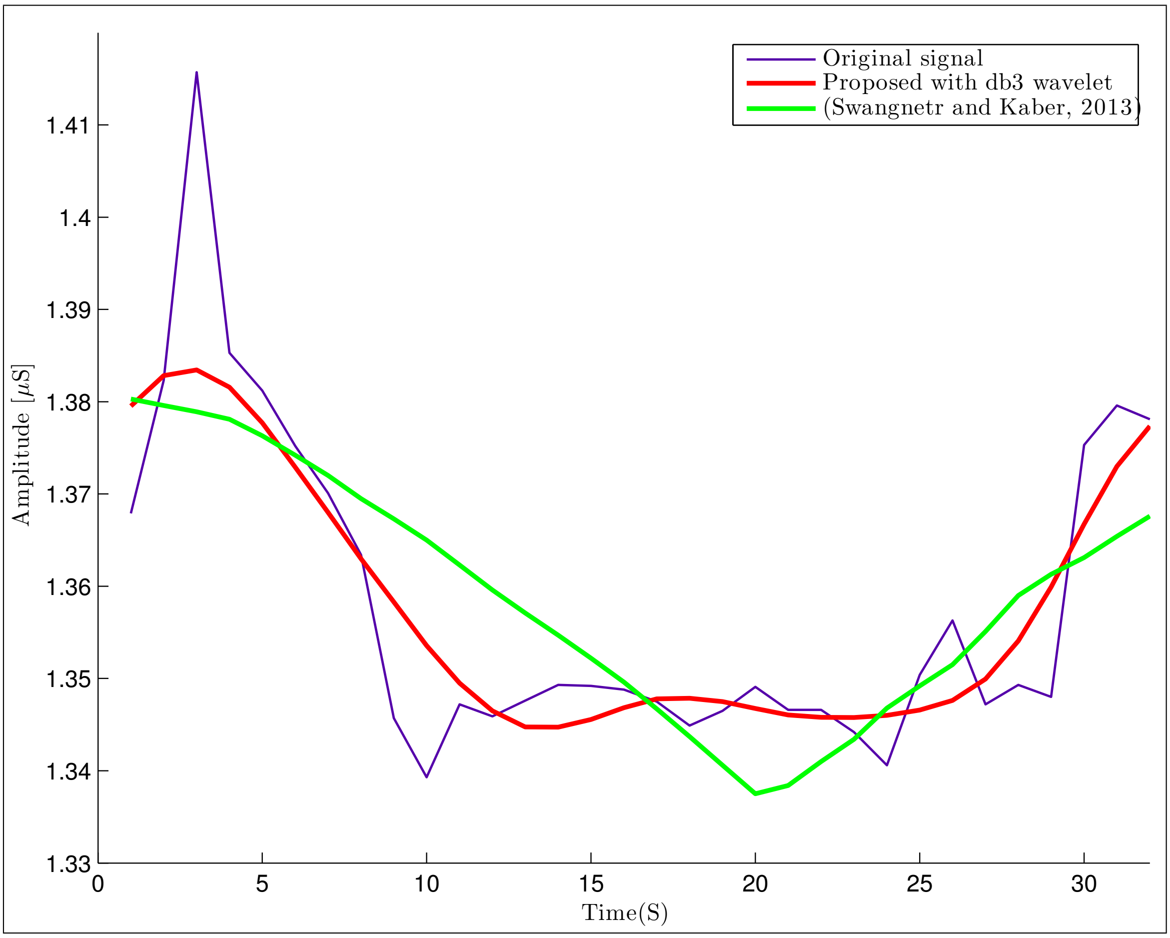


Fig. 7.4 Original EDA and denoised signals processed by proposed method with *db*3 and method proposed by (Swangnetr and Kaber, 2013)

artifact power reduction (at a cost of a higher distortion of the artifact-free signals). Thus, the performance of the proposed methods using the *coiflet*3 and *haar* method were separately compared to the performance of the (Chen et al., 2015) method which is using the *haar* wavelet. The median values obtained for the APA and NMSE metric for each of these three methods are given in table 7.4.

The *t*-test comparisons (table 7.5) showed that there were statistically significant differences between all the three methods both, in terms of the APA and the NMSE. The values suggest that the two versions of the proposed method outperform the method proposed by (Chen et al., 2015) in terms of NMSE, that is, the proposed methods with *coiflet*3 and *haar* variants introduce significantly less distortion in the artifact-free signal. By contrast, the proposed methods obtained lower values in the APA metric compared to the (Chen et al., 2015) method, that is they reduce less the power of the artifacts. However, table 7.1 shows that the APA values of both *haar* variant of the proposed method and (Chen et al., 2015)

Table 7.4 Median of NMSE and APA for evaluated methods

Method	*I*	*II*	*III*
APA (dB)	1.059	3.233	3.695
NMSE (dB)	-18.170	-13.756	-6.443

I–$Coiflet3$ *II*–$Haar$ *III*–(Chen et al., 2015)

Table 7.5 t-test of NMSE and APA for evaluated methods

	APA			**NMSE**		
	t	p	d	t	p	d
Coif3-haar	12.96	<.001	1.58	3.92	<.001	0.68
Coif3-*I*	17.74	<.001	2.16	15.58	<.001	2.71
Haar-*I*	3.54	<.001	0.43	7.79	<.001	1.36

I–(Chen et al., 2015)

methods are close, and the difference, although statistically significant, represents a small (0.43) effect size (Cohen, 1988), while the difference in NMSE between these two methods has a large effect size (1.36). Indeed, when compared to the (Chen et al., 2015) method the *haar* version of the proposed method seems to introduce much lesser distortion in the artifact-free signal, while reducing the power of the artifacts almost as well as the (Chen et al., 2015) method.

Summarizing, the version of the proposed method that uses a mother wavelet fitting the signal waveform (*db*3) equals the state-of-the-art method (Swangnetr and Kaber, 2013) using the *db*3 basis function, in terms of artifact power reduction, and taken together, these results suggest that the *db*3 version equals the (Swangnetr and Kaber, 2013) method in terms of reduction of noise and shows an advantage compared to it by introducing less distortion in the artifact-free signals. At the same time, the *haar* version of the proposed method, based on a mother wavelet fitting the artifacts waveform, is better than the state-of-the-art method based on this approach ((Chen et al., 2015)), in terms of distortion of the artifact-free signals, and almost equals it in terms of artifact power reduction. Hence, the next question to address is which one of these two versions shows an overall better performance. Two t-tests were run comparing the APA and NMSE values provided by the *db*3 and *haar* variant of the proposed methods. The results evidence that the difference between the two methods was statistically significant both in terms of the APA ($t(66) = 10.56; p < 001$) and NMSE ($t(32) = 2.97; p = .006$). However, the fact that the effect size was much larger in the case of APA (Cohen's $d = 1.29$) than in the case of the NMSE metric (Cohen's $d = 0.51$) suggests

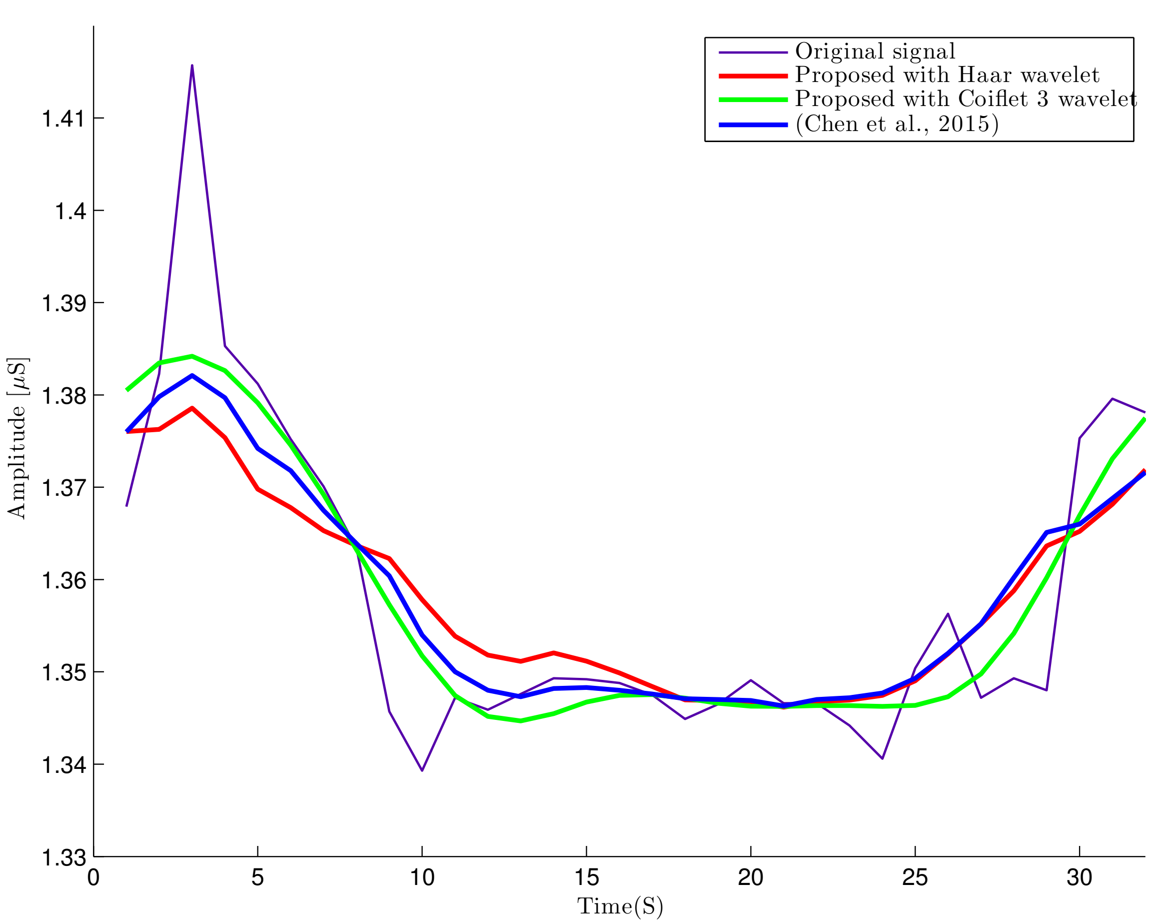


Fig. 7.5 Original EDA and denoised signals processed by proposed method with $Coiflet3$, proposed method with $Haar$ and method proposed by (Chen et al., 2015)

that the advantage of the *haar* variant method in reduction of the artifact power is stronger than the advantage of the $db3$ variant method in NMSE for the introduction of distortions; so indicating that the *haar* method is the best one among the three versions ($db3$, *haar* and the $coiflet3$) of the proposed method.

7.5 Online Validation : MIST Experiments

In order to test the performance of the algorithm in actual online tasks, we asked 12 volunteers to participate in an experiment aimed to elicit different levels of physiological arousal on them. The goal of this experiment was to analyze whether the online filtering of EDA signal done with our *haar* method improves the identification of SCRs, and, ultimately, allows a better detection of cognitive and emotional states. SCRs are one of the most well-known features of EDA activity, consisting of momentary increases in the EDA level occurring when the

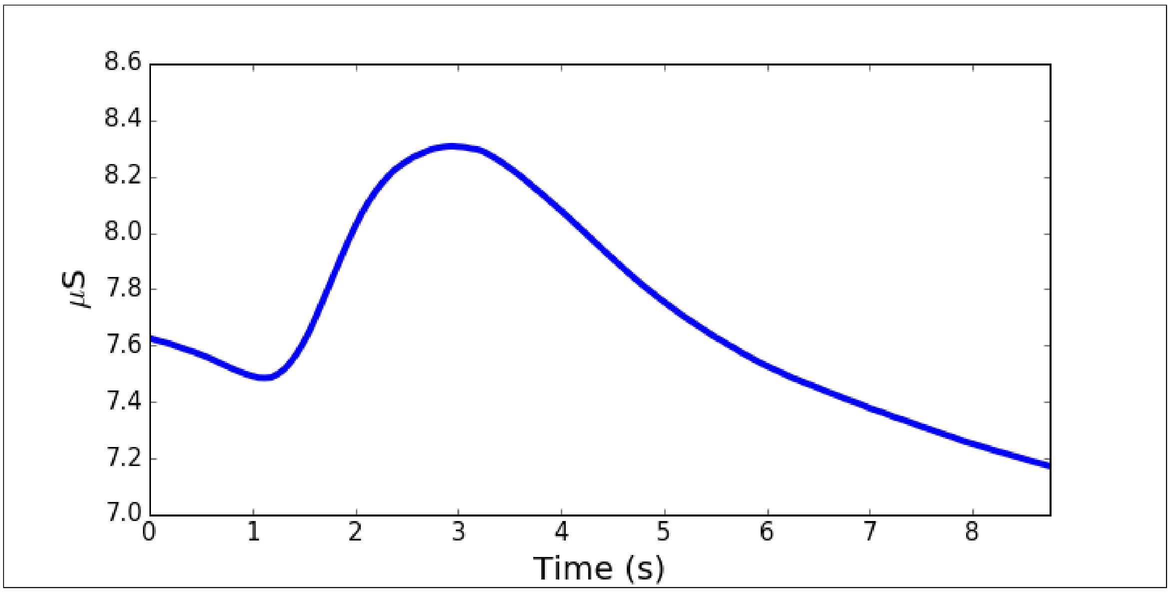


Fig. 7.6 Example SCR

individual encounters a stimulus eliciting arousal in the sympathetic nervous system. SCRs commonly start between 1 and 3 seconds after stimulus presentation, and are characterized by a clearly identifiable waveform (see figure 7.6), which includes a sudden rise in EDA levels, with an amplitude typically between 0.1 and 1.0 μS, and a rise time between the onset of the response and the peak between 1 and 3 seconds (Dawson et al., 2007).

The experiment involved three tasks aimed, respectively, to induce states of relax, moderate arousal, and stress, in the participants. The logic behind this experimental design is that, since SCRs are much more frequent when the individual is aroused (Boucsein, 2012), a higher number of SCR should be found for the more stressful tasks compared to the more relaxing tasks.We implemented the tasks using the Psychopy software[2].

7.5.1 Tasks Description

In the first task (*relax* task), the participant was instructed to watch a series of pictures of landscapes displayed on a PC monitor while listening to a segment of calm classical music. Each picture was shown during two seconds, and the total duration of the task was two minutes.

In turn, the second and third tasks consisted of an adapted version of the Montreal Imaging Stress Task (MIST) (Dedovic et al., 2005). It has two parts: in the first one (*training* task), the participant was asked to solve a series of arithmetic operations with one or two

[2]http://www.psychopy.org/

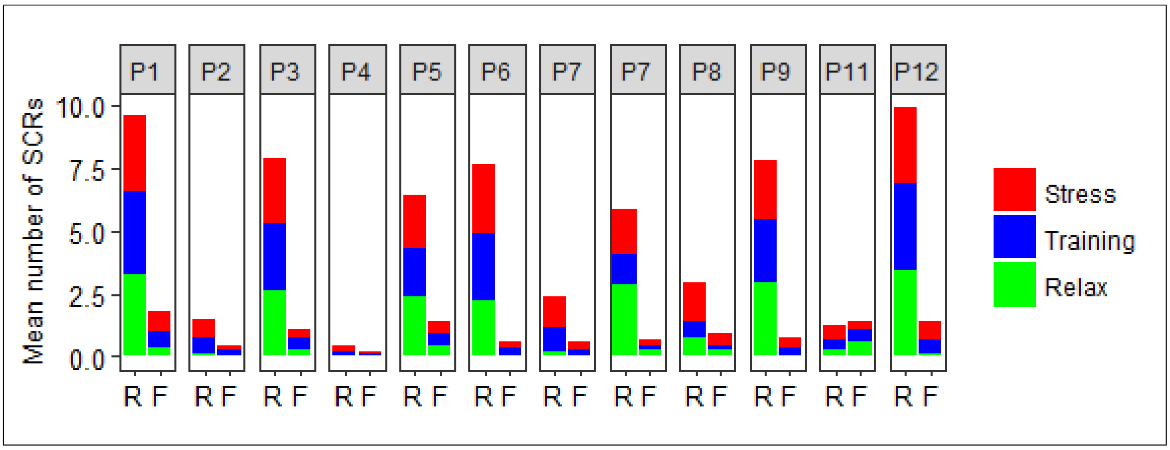


Fig. 7.7 Mean number of SCRs across participants by task for raw and filtered signal, P*i*: Participant *i*, R: Raw Signal, F: Filtered Signal

digits in the shortest possible time, but without any time limit. This task was expected to elicit moderate states of arousal and served to gather an approximation of the time that the participant needed for solving such operations. The second part of the test (*stress* task) presented similar operations but with a time limit, which was shown in a time bar on the screen. The software calculated the mean time required by the participant for each operation from the first part, and assigned a 10% less time for each operation in the second part, making it very hard to solve the operations correctly. The program provided feedback on whether the responses were correct or not, and also adjusted the time assigned to each operation, in a way such, after three correct responses, the time assigned was reduced another 10% , so making the participant unable to solve correctly the operations. In order to avoid excessive frustration in the participant, after three wrong responses, the time assigned to the next operation was slightly increased, so the participant had more chances to give a correct answer. The *stress* task also included negative feedback by displaying on the screen a bar showing a (fake) comparison between the participant's performance and the mean population performance in the task, which indicated that the participant's performance was below the mean. After several operations with incorrect responses, a warning was presented on the screen indicating that the poor performance of the participant was compromising the results, and requested to invest more effort in the task. Since temporal pressure and negative feedback are two of the most common causes of stress, this task is expected to elicit higher stress compared to the training task. The training and stress tasks required the participants to manually type the responses on the keyboard, which involved moving the right hand to which the electrodes were attached, thus favoring the occurrence of motion artifacts in the EDA signal.

Fig. 7.8 Experimental set-up

7.5.2 Procedure

Twelve employees (six males and six females) from Instituto de Robótica Para La Dependencia (IRD), Barcelona, Spain, aged between 21 and 54 (M = 41.92, SD = 8.43), and unaware of the experiment's purposes, voluntarily participated in the trial. Each one carried out the experimental tasks while his or her EDA signals were recorded and filtered online following the method described above using the *haar* mother wavelet. Figure 7.8 shows the experimental setup.

During the experiment, participants were comfortably seated while wearing a noise canceling headphone in an experiment room. The experiment was presented on a computer screen while participants EDA was collected using a wireless wristband sensor Shimmer GSR+[3]. A laptop was placed behind the intervention table, hidden from the participant. It was used to present the MIST on the monitor, to receive the data from the Shimmer GSR+ sensor and to denoise and save the denoised EDA data.

[3] http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

The whole session lasted for around 15 minutes. To achieve an efficient time synchronization, the experiment execution and the EDA collection and filtering were done on the same system and were timestamped with the system time. For online filtering, the proportion of motion artifacts in the original signal, $\delta = 0.10$ was employed for all participants. The data collection window was of 8 seconds duration after which the filtering was applied and the data was saved in a new file. A sliding window protocol of 1 second was applied. The original and the denoised data corresponding to every single window was saved in a separate file for the ease of further processing in validation of the denoising results. After finishing the tasks, participants were asked to rate on a 7-point Likert-type scale how stressful each task was, in order to have a measurement of the subjectively perceived stress. Finally, participants were debriefed about the goals of the experiment and the manipulation check in the stressful task.

For the validation of the denoised results, an SCR detector was implemented in Matlab. SCR detection was a two-stage process. First, a valley is located on the signal. Then the highest peak after the valley is found. The SCR amplitude is the highest peak amplitude minus the lowest valley amplitude. SCR occurrences that were not large enough ($< 0.1 \mu S$) were rejected.

7.5.3 Results

According to the self-reported estimates provided by the participants, the three tasks achieved their goals of eliciting different states of subjective stress. The mean rating for the relax task in the stress scale was 1.42 (*SD* = 1.16), while the mean rating for the training task was 3.00 (*SD* = 1.76), and the mean rating for the stress task was 5.83 (*SD* = 1.34), which, in a scale ranging from 1 to 7, can be considered as low, moderate, and high values, respectively.

Since the duration of the tasks was different, we consider the mean number of SCRs by 8 seconds segment in each task instead of the absolute number of SCRs by the task. The first noticeable aspect of the comparison between the raw and the filtered signal is that the mean number of SCRs is much higher in the raw signals than in the filtered ones (see Table 7.6). The values for the raw signals indicate that, in some cases, more than one (even up to two or

Table 7.6 Mean (and Standard Deviation) of mean number of SCRs per segment across participants by task, as calculated from the raw and the filtered signal

Task	**Relax**	**Training**	**Stress**
Raw Signal	1.75 (1.36)	1.72 (1.19)	1.82 (0.97)
Filtered Signal	0.22 (0.20)	0.34 (0.17)	0.39 (0.22)

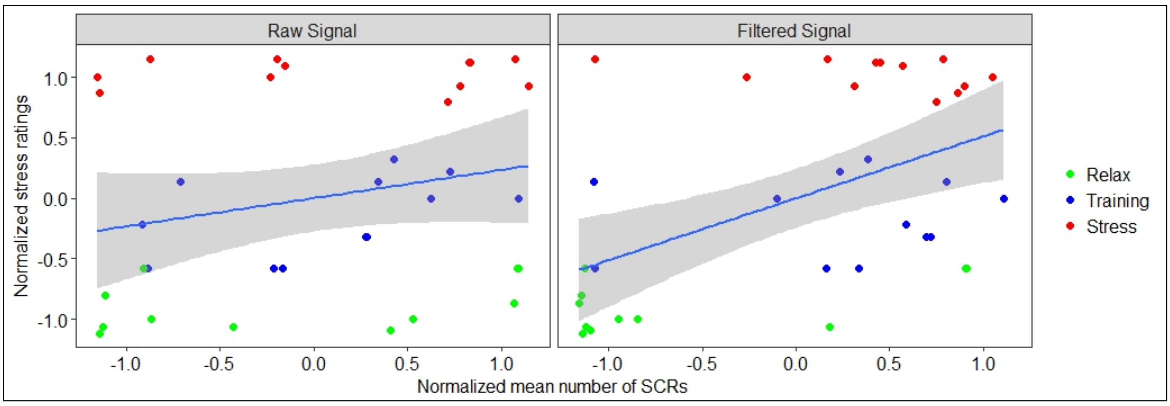


Fig. 7.9 Linear Regression of SCRs by task predicted the normalized stress ratings

three) SCRs were detected by each 8-seconds segment, even for the relaxing task (figure 7.7). Since these values deviate largely from the typical number of SCRs (Dawson et al., 2007), they are not realistic and probably a consequence of counting motion artifacts as SCRs. By contrast, the filtered signal shows much more realistic values (less than one SCR by segment in all cases). Particularly for the relax task, where few or none SCRs were expected, the raw signal shows a considerably high number of them, while they are kept to a minimum in most cases when using the filtered signal (see figure 7.7). Meanwhile, the training and stress task show a higher number of SCRs, which is in agreement of what can be expected from this test. Taken together, these results suggest that a number of artifacts were counted as SCRs when using the non-filtered signal and that the filtering process helped to solve this problem.

In order to check whether the filtering improves the detection of SCRs, we analyzed the correlation between the number of SCRs in each task and the subjective stress scores given by the participants. Since the presence of SCRs shows a high individual variability, for analysis we used within-subject normalized scores (Jennings and Gianaros, 2007) of mean number of SCRs by segment, for both the measurements with the raw and the filtered signal. In order to correct for possible different baselines in subjective perception of stress, the stress ratings were also normalized within-subjects. Then a linear regression was fitted to explore how the normalized mean number of SCRs by task predicted the normalized stress ratings, for both the raw and the filtered signal (figure 7.9). The results for the raw signal show no significant relationship between the number of SCRs and the stress ratings ($\beta = 0.23, p = 0.17$), while, in the case of the filtered signal, the number of SCRs is a significant predictor of subjective stress ($\beta = 0.51, p = 0.001$). As shown in figure 7.9, while there is no clear pattern of relationship between SCRs count and stress ratings when using the raw signal, in the data obtained with the filtered signal, even though the presence of some outliers, the relationship is more linear. Moreover, while in the case of the raw signal the variance in the dependent variable explained by the model is quite low ($R^2 = 0.05, F(1,34) = 1.96, p = 0.17$), in the regression considering the filtered signal it explains about a quarter of the variance in the dependent variable ($R^2 = 0.26, F(1,34) = 12.12, p = .001$), that is, five times more variance in subjective ratings is explained by SCRs count when using the filtered signal. Noticeably, the data obtained with the raw signal show a relationship between SCRs count and which has an effect size in the range small to medium ($b = 0.23$) according to (Cohen, 1988), and which is not statistically significant. It would lead a researcher to interpret that there is no statistical evidence of a relationship between both variables. By contrast, the results obtained with the filtered signal shows a large to very large effect size ($\beta = 0.51$) (Cohen, 1988), whose statistical significance is very high ($p = .001$), so there is strong evidence of a clear relationship between subjective stress ratings and number of SCRs.

These results suggest that the unfiltered signal contained a significant number of motion artifacts that were mistaken as SCRs by the detector. On the other hand, the online filtering allowed cleaning such artifacts, thereby improving the detection task and making the signal more descriptive of the actual subjective stress experienced by the participants.

7.6 Discussion and Conclusion

The results demonstrate the effectiveness of the proposed method for online EDA filtering in real-world scenarios. Specifically, *haar* variant of the proposed method performs an artifact power reduction similar to the currently available methods, with a higher respect for the original artifact-free signals. Furthermore, this is achieved with much lower computational cost than the state-of-the-art methods. In addition, the proposed method does not have any dependency on external libraries such as QP Solver required by (Greco et al., 2016) or Expectation-Maximization (EM) algorithm required by Chen et al. (2015) approach. It is worthwhile noting that the cvxEDA approach by Greco et al. (2016) requires tuning of at-least four parameters (τ_0 slow time constant, τ_1 fast time constant, α penalization for the sparse SMNA driver and γ penalization for the tonic spline coefficients) in pre-processing step. On the other hand, the proposed method requires knowledge of a single parameter (the proportion of motion artifacts in original signal δ) only in the pre-processing step.

We envision that future affective computing applications will be based on wearable devices, which are characterized by limited computational power and storage. Hence, the lower computational requirements of the proposed method along with no external dependencies can facilitate the embedding of denoising algorithms into such devices. It can, therefore, boost the creation of applications able to monitor user's cognitive and emotional states. In this sense, stress prediction from EDA signal is still an open research issue, which falls far beyond our objectives in this work, and which probably will require the use of more significant features than the one used here. Even so, our results confirm that the online filtering significantly improves the quality of the signal in such a way that the correlation between the analyzed signal feature (SCRs) and subjectively perceived stress is substantially increased.

One limitation of the present study is that the performance of our filtering algorithm was tested in three contexts in which the users carried out actions (driving dataset, cognitive stimulation sessions and typing in the keyboard) involving hand and arm movements while seated. However, wearable devices allow EDA recording in more dynamic situations, like walking or practicing sports, which can involve larger and faster movements in the subject's body and arms and that may, therefore. produce more frequent or larger motion artifacts. Whereas testing the algorithm in all possible contexts of use of EDA monitoring is virtually impossible, further research should validate the effectiveness of the proposed method in some representative situations (e.g. customers walking in a grocery store, for neuromarketing studies). However, since many of the possible applications of EDA monitoring (including virtually in any ICT system addressed to recognize emotions and engagement level from physiological signals, such as tactile gaming consoles, robots during the human-robot interaction, neuro-marketing, etc.) involve users in more static situations like the ones analyzed in the present study, the results of the present research mean a valuable contribution even without further validation in more dynamic situations.

Finally, there are reasons to think that the presence of motion in EDA signal cannot only be seen as a source of a noise. Since it has been found that humans move or gesture in response to affective changes (Chellali and Hennig, 2013), the temporal concurrence of EDA changes and movements could mean a meaningful source of information on user's cognitive and emotional processes. For instance, a research question could be whether the levels of movement associated with EDA changes are informative of other processes related to physiological arousal (e.g. strategies for emotion regulation or coping with high cognitive load or stress). Since wavelet-based motion artifact removal for EDA may also have advantages in this sense, the present research paves the way for future research addressing this issue.

Chapter 8

Feature Extraction and Selection for Emotion Recognition from EDA

Summary. Many methods for EDA feature extraction have been studied; however, their suitability for emotion recognition has been tested using a small number of distinct feature sets and on different, usually small data sets. A major limitation is that no systematic comparison of EDA features exists. We reviewed feature extraction methods for emotion recognition from EDA based on 25 studies. We compared these features for feature selection using machine learning techniques on a publicly available AMIGOS dataset. We present the results of the performance of three feature selection methods and usage of selected feature types across time, frequency and time-frequency domains. We did not find any statistical evidence that any of the three employed feature selection methods outperform the others. However, the subject-dependent classification results were significantly higher than the subject-independent classification for both the arousal and the valence recognition. MFCC and related statistical features were explored for the first time for the emotion recognition from EDA signals and they outperformed all other feature types, including the most commonly used SCR related features. We also compared our results with methods employed by researchers of AMIGOS dataset for classification of emotional states and they show that the EDA features explored in this study provided better performance, validating the findings of our study.

It is not known which features are most appropriate in emotion recognition from EDA and to the best knowledge of the author, none of the previous works has made any systematic comparison of EDA features. Hence, we aim to provide a complete review of feature extraction methods used for emotion recognition from EDA signals and a systematic comparison of reviewed EDA features on a publicly available dataset to identify significant EDA features for affective analysis. We reviewed feature extraction techniques from 25 studies and then implemented and evaluated them using three feature selection methods on a publicly available dataset of annotated EDA signals from 37 participants. Hence, the contributions of this work are as follows:

1. We provide an inclusive review of EDA features for emotion recognition.
2. We provide a first-ever systematic comparison of features on one database using multiple feature selection methods.
3. We identify the most significant EDA features for emotion recognition.

The remainder of this chapter is structured as follows: Section 8.1 reviews feature extraction methods used for emotion classification from EDA. Three different state-of-the-art feature selection techniques are explained in Section 8.2. In Section 8.3, we present details of the employed public dataset and the processing of EDA signal for feature comparison. We provide the results in Section 8.4 with its discussion in Section 8.5. Finally, Section 8.6 concludes the chapter by summarizing this work and highlighting future steps.

8.1 Feature Extraction

In this section, we review a wide range of features relevant for emotion recognition from EDA that have been proposed in the past. Based on similar work on EEG signals by (Jenke et al., 2014), we distinguish features in time domain, frequency domain, and time-frequency domain. We compiled a comprehensive feature list for emotion recognition from EDA based on 25 studies and the feature list is presented in table 8.1. The references used to compile the EDA feature list is presented in table 8.2.

Table 8.1 EDA features used in the research

Domain	Features	Parameters	Description
Time Domain	SCR Features	peakCount	SCR amplitude peak counts
		meanpeakAmplitude	SCR amplitude mean
		Mean Rise Time	SCR amplitude mean rise time
		Sum Peak Amplitude	SCR amplitude summation
		Sum Rise Time	SCR amplitude rise time summation
		Sum Areas	Sum of estimated areas of orienting responses
	Statistical Features	auc	Area under curve
		meanEDA	Mean of signal
		stdEDA	Standard deviation of signal
		kurtEDA	Kurtosis of signal
		skewEDA	Skewness of signal
		meanDerivative	Mean of 1st order derivatives
		meanNegativeDerivative	Mean of negative values of 1st order derivatives
	Hjorth Features*	Activity	Variance of the signal
		Mobility	Square root of variance of the first derivative of the signal divided by variance of the signal
		Complexity	Mobility of Mobility
	Higher Order Crossings*	HOC	Sequence of zero-crossings of a specific sequence of filtered signal
Frequency Domain	Statistical Features	SMA	Signal Magnitude Area
		meanEDA	Mean of signal

Frequency Domain	Statistical Features	stdEDA	Standard deviation of signal
		signalRange	Range of Signal
		harmonicsSummation	Summation of FFT Harmonics
		kurtEDA	Kurtosis of signal
		skewEDA	Skewness of signal
		signalEnergy	Energy of the signal
	Band Power	SpectralPower [0.5-0.5 Hz], $\delta = 0.1$	5 spectral power in the [0-0.5]Hz bands
		minSpectralPower	Minimum of spectral band powers
		maxSpectralPower	Maximum of spectral band powers
		varSpectralPower	Variance of spectral band powers
Time-Frequency Domain	Wavelet Features	Energy distribution	Energy percentage for wavelet levels
		energyWavelet	Energy for wavelet levels [Absolute]
		entropyWavelet	Entropy for wavelet levels
		rmsWavelet	Root mean square for wavelet coefficients
		coefficients	Approximation and detail coefficients
	MFCC*	mfccCoefficients	MFCC coeffcients for all frames
		meanMFCC	Mean of MFCC coefficients
		stdMFCC	Standard deviation of MFCC coefficients
		medianMFCC	Median of MFCC coefficients
		kurtMFCC	Kurtosis of MFCC coefficients
		skewMFCC	Skewness of MFCC coefficients

8.1.1 Time Domain Features

Given the non-stationary nature of EDA, researchers on the field of psychophysiology using EDA have traditionally focused on time-domain features of the EDA signals.

SCR Features

In certain cases researchers have focused on event-related features of EDA, that is attributes of the short-term response a few seconds after the presentation of a certain stimulus (i.e. images or sounds), such as the presence or absence of a SCR. This can be faced based on an event-related analysis, that is, the detection of event-related features of EDA, such as SCRs, in a short time windows after stimulus presentation, which, although useful in laboratory research contexts, is not applicable to real-life contents. Alternatively, phasic responses of EDA, such as SCRs, can be automatically detected in longer time windows and features can be extracted from them. This involves the definition of thresholds, in order to restrict the analysis only to non-negligible responses, and discarding those small changes in the signal that do not reach the thresholds and therefore cannot be considered as SCRs. In this sense, a traditional threshold for SCR amplitude has been $0.05\mu S$ (Boucsein, 2012).

In the present work, the following event-related features are considered to characterize EDA time series: SCR Amplitude (Giakoumis et al., 2011; Greco et al., 2017; Kim and André, 2008; Koelstra et al., 2012; O'Connell et al., 2008; Piacentini, 2004; Sano and Picard, 2013; Zhai et al., 2005), SCR peak count (Giakoumis et al., 2011; Greco et al., 2017; Healey and Picard, 2005; Kim and André, 2008), mean SCR amplitude (Healey and Picard, 2005; Piacentini, 2004; Sano and Picard, 2013), mean SCR rise time (Giakoumis et al., 2011; Healey and Picard, 2005; Piacentini, 2004; Zhai et al., 2005), sum of SCR peaks amplitudes (Healey and Picard, 2005; Piacentini, 2004; Sano and Picard, 2013), sum of SCR rise times (Healey and Picard, 2005; Koelstra et al., 2012; Piacentini, 2004; Zhai et al., 2005), area under the curve of SCRs (Greco et al., 2017), and sum of SCR areas (Barreto et al., 2007; Greco et al., 2017; Healey and Picard, 2005).

Statistical Features

Some of the most commonly used time-domain features refers to statistical parameters of the signal during a relatively long recording, which are useful when the stimulus of interest is not limited to a very specific moment but develops over time (e.g. when watching a movie, listening to a song, browsing a website, or giving a speech). Previous research has considered several statistical measures of time domain representation of the EDA signal (Jenke et al., 2014):

Table 8.2 References used to compile EDA Features

Serial Number	**Reference**
1	(Hjorth, 1970)
2	(Davis and Mermelstein, 1980)
3	(Yakowitz, 1994)
4	(Piacentini, 2004)
5	(Healey and Picard, 2005)
6	(Zhai et al., 2005)
7	(Barreto et al., 2007)
8	(Wang and Gong, 2008)
9	(Horlings et al., 2008)
10	(Kim and André, 2008)
11	(O'Connell et al., 2008)
12	(Petrantonakis and Hadjileontiadis, 2010)
13	(Hosseini and Khalilzadeh, 2010)
14	(Giakoumis et al., 2011)
15	(Koelstra et al., 2012)
16	(Sano and Picard, 2013)
17	(Höller et al., 2013)
18	(Kurniawan et al., 2013)
19	(Giakoumis et al., 2013)
20	(Leite et al., 2013)
21	(der Zwaag et al., 2013)
22	(Swangnetr and Kaber, 2013)
23	(Jenke et al., 2014)
24	(Ghaderyan and Abbasi, 2016)
25	(Greco et al., 2017)

- Power

$$P_X = \frac{1}{N}\sum_{-\infty}^{\infty}|X(n)|^2 \tag{8.1}$$

- Mean (der Zwaag et al., 2013; Ghaderyan and Abbasi, 2016; Giakoumis et al., 2011; Greco et al., 2017; Healey and Picard, 2005; Hosseini and Khalilzadeh, 2010; Kurniawan et al., 2013; Leite et al., 2013; Piacentini, 2004; Wang and Gong, 2008)

$$\mu_X = \frac{1}{N}\sum_{n=1}^{N} X(n) \tag{8.2}$$

- Standard deviation (Ghaderyan and Abbasi, 2016; Giakoumis et al., 2011; Greco et al., 2017; Healey and Picard, 2005; Hosseini and Khalilzadeh, 2010; Kurniawan et al., 2013; Leite et al., 2013; Piacentini, 2004)

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\mu_X)^2} \tag{8.3}$$

- Kurtosis (Ghaderyan and Abbasi, 2016; Giakoumis et al., 2013; Hosseini and Khalilzadeh, 2010; Leite et al., 2013)

$$\frac{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\mu_X)^4}{\left[\sqrt{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\mu_X)^2}\right]^2} \tag{8.4}$$

- Skewness (Ghaderyan and Abbasi, 2016; Giakoumis et al., 2013; Hosseini and Khalilzadeh, 2010)

$$\frac{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\mu_X)^3}{\left[\sqrt{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\mu_X)^2}\right]^3} \tag{8.5}$$

- Mean of the 1st difference (Giakoumis et al., 2011; Hosseini and Khalilzadeh, 2010; Kim and André, 2008; Wang and Gong, 2008)

$$\overline{\delta}_X = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\overline{X}(n+1)-\overline{X}(n)\right| = \frac{\delta_X}{\sigma_X} \tag{8.6}$$

- Mean of the 2nd difference (Hosseini and Khalilzadeh, 2010; Koelstra et al., 2012)

$$\overline{\gamma}_X = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|\overline{X}(n+2)-\overline{X}(n)\right| = \frac{\gamma_X}{\sigma_X} \tag{8.7}$$

Hjorth Features

For a given time series signal of the EEG signal, following Hjorth features can be calculated (Hjorth, 1970):

- Activity:
$$A_X = \sum_{n=1}^{N} (X(t) - \mu)^2 T \tag{8.8}$$
- Mobility:
$$M_X = \sqrt{var(X(\dot{n}))var(X(n))} \tag{8.9}$$
- Complexity:
$$C_X = M(X(\dot{n}))M(X(n)) \tag{8.10}$$

Hjorth features have been used in several EEG related studies (Horlings et al., 2008; Höller et al., 2013) and it motivated us to explore its nature for emotion recognition on EDA signals.

Higher Order Crossings

We also explored the feature set drawn from higher order crossings (HOC) analysis of EDA time-series signals (Yakowitz, 1994). When a specific sequence of filters is applied iteratively to a time series Z_t, the corresponding sequence of zero-crossings is known as HOC sequence and is represented by:

$$\Im_k\{Z_t\} = \nabla^{k-1} Z_t \tag{8.11}$$

∇ is iteratively applied backward difference operator $\nabla Z_t \equiv Z_t - Z_{t-1}$ and the order k is selected from $1.....,50$ that would result in the maximum classification rate of the given signals (Petrantonakis and Hadjileontiadis, 2010). Desired simple HOC are then obtained by counting the symbol changes D_k in $\Im_k\{Z_t\}$. HOCs are then used to construct the feature vector $\mathrm{FV_{HOC}}$ as follows:

$$\mathrm{FV_{HOC}} = [D_1, D_2, \ldots, D_L], \quad 1 < L \leq J \tag{8.12}$$

HOC features have been found robust and consist to effectively discriminate emotions from other physiological signals (Jenke et al., 2014; Petrantonakis and Hadjileontiadis, 2010) and hence we decided to investigate it in this research.

8.1.2 Frequency Domain Features

The transient characteristics of the EDA signal can be understood better by analyzing the frequency domain representation of the signal. The frequency domain analysis has shown superior capability for the gradient component's detection of individual SCR over traditional amplitude analysis (Shimomura et al., 2008). Fast fourier transform (FFT), short-time fourier transform (STFT) and power spectra density (PSD) estimation using Welch's method are the most commonly used algorithms to obtain the frequency domain representation of the signal.

Due to the different rate of physiological processes, EDA signals vary significantly with the frequency (Ghaderyan and Abbasi, 2016). Hence, frequency oscillations of EDA signals can be divided into different frequency subbands to analyze it more detailed manner. Similar analysis has been performed for other physiological signals such as EEG and EMG (Fallahi et al., 2016; Wilson and Fisher, 1995). The recommended frequency range of EDA signal ($0.5 - 0.50Hz$) was split in five bands following the suggestions of the previous literature (Ghaderyan and Abbasi, 2016; Wang and Gong, 2008). The features extracted from the resulting frequency domain representation of the EDA signal are: a set of statistical features (variance, range, signal magnitude area, skewness, kurtosis, harmonics summation) and spectrum power of five frequency bands, their minimum, maximum, and variance (Alberdi et al., 2016; Ghaderyan and Abbasi, 2016; Koelstra et al., 2012; Wang and Gong, 2008).

8.1.3 Time-Frequency Domain Features: Wavelets

Since recorded EDA data exhibits non-stationary behavior, hence wavelets have been found suitable for modeling EDA activity.

Discrete Wavelet Transform

When the wavelets are discretely sampled, the wavelet transform is known as Discrete Wavelet Transform (DWT). Denoised DWT wavelet coefficients have been used as EDA features for emotional state classification (Swangnetr and Kaber, 2013). Wavelet analysis of a signal consists of translations ($k \in \mathbb{Z}$) of the *father wavelet* $\phi(t)$ *(scaling function)* and dilations and translations ($k \in \mathbb{Z}$, $j \in \mathbb{Z}$) of the *mother wavelet* $\psi(t)$. The wavelet series representation of a signal $x(t)$ is then:

$$x(t) = \sum_j c_{j,k}\phi_k(t) + \sum_j \sum_k d_{j,k}\psi_{j,k}(t) \tag{8.13}$$

where $(c_{j,k})$ are the *approximation coefficients* and $(d_{j,k})$ are the *detail coefficients* of the wavelet coefficient set and are calculated as follows:

$$c_{j,k} = \langle x(t), \phi_{j,k}(t) \rangle \tag{8.14}$$

$$d_{j,k} = \langle x(t), \psi_{j,k}(t) \rangle \tag{8.15}$$

In DWT, wavelet acts as a bandpass filter from signal processing point of view, where the scaling and wavelet functions serve as low pass (h[n], Equation 8.16) and high pass filters (g[n], Equation 8.17), respectively.

$$\phi(t) = \sum_n h[n]\sqrt{2}\phi(2t-n) \tag{8.16}$$

$$\psi(t) = \sum_n g[n]\sqrt{2}\psi(2t-n) \tag{8.17}$$

where,

$$g[n] = h[2N-1-n] \tag{8.18}$$

When applied, above decomposition halves the time resolution and doubles the frequency resolution. Above procedure can be applied iteratively for multilevel decomposition of the signal. Wavelet decomposition levels corresponds to different frequency bands and this correspondence is based on the sampling frequency of the signal.

Stationary Wavelet Transform

Stationary Wavelet Transform (SWT) is redundant, linear and hence shift invariant in comparison to the DWT (Nason and Silverman, 1995). SWT also provides better sampling rates in the low frequency bands compared with a standard DWT (Nason and Silverman, 1995). SWT technique has been successfully applied to de-noise EDA signals with high efficacy and less computational complexity (Shukla et al., 2018b). Therefore, we extracted following wavelet features based on SWT:

- Wavelet Energy: $E(j) = \sum_{i=1}^{j} D_i(n)^2$
- Wavelet Decomposition Energy: The percentage of energy corresponding to the approximation and to the details coefficients.

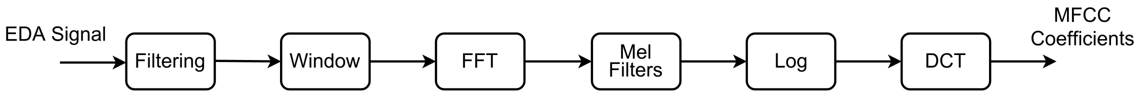


Fig. 8.1 MFCC Feature Extraction

- Wavelet Entropy: $E(j) = -\sum_{i=1}^{j} D_i(n)^2 log(D_i(n)^2)$
- Wavelet Root Mean Square (RMS): $RMS(j) = \sqrt{\frac{1}{N}\sum_{i=1}^{j} |D_i(n)|^2}$

8.1.4 Mel-Frequency Cepstrum

The change in the skin conductance (SC) of the EDA signal can be characterized by a sequence of overlapping fast vary- ing phasic skin conductance responses (SCRs) overlying a slowly varying tonic activity (i.e., skin conductance level; SCL). This superposition complicates the proper decomposition of the SC data and hence limits the ability of classical methods for the assessment of SC responses (Boucsein, 2012). Sudomotor nerve activity (SMNA) can be considered as a driver and it consists of a sequence of mostly distinct impulses (i.e., sudomotor nerve bursts). These bursts trigger the specific impulse response (i.e., SCRs) and thereby the SC can be modeled by driver-impulse response (IR) convolution. As a result of this process, SC can be represented by: (Benedek and Kaernbach, 2010b)

$$\begin{aligned} SC &= SC_{tonic} + SC_{phasic} \\ &= (Driver_{tonic} + Driver_{phasic}) * IR \end{aligned} \tag{8.19}$$

In this model, SC is considered as the output of the skin system driven by an excitation sequence of sudomotor nerve bursts. But such convolution of the response and the drivers can not be easily separated in time domain. Cepstrum analysis (CA) is an important technique for analyzing similar models of the speech signal (Benesty et al., 2007).

Cepstrum of a discrete-time signal is the inverse discrete-time fourier transform (IDTFT) of the logarithm of the magnitude of the discrete-time Fourier transform (DTFT) of the signal and is given by:

$$c[n] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \log|X(e^{i\omega})| e^{i\omega n} d\omega \tag{8.20}$$

where $X(e^{i\omega})$ is the DTFT of the signal.

CA has been used successfully to isolate the basic waveform and the excitation function of the physiological signals such as electrocardiogram (Li and Narayanan, 2010), electroencephalogram (Kamath, 2013) and even EDA (Ghaderyan and Abbasi, 2016). While analyzing EDA signals using CA, it was concluded that CA could be useful for analysis of superimposed EDA signals given its ability to magnify small amplitude variations (Ghaderyan and Abbasi, 2016). It motivated us to further explore Mel-frequency cepstral coefficients (MFCC) as a feature vector of EDA signals. MFCC is a new type of cepstrum representation based upon the weighted cepstrum distance measures (Davis and Mermelstein, 1980) and has become widely established for many pattern recognition problems related to speech signals (Benesty et al., 2007).

The process of extracting MFCC features as applied to EDA signals is shown in figure 8.1 and is explained below:

1. The EDA signal is filtered to remove the motion artifacts using the sophisticated SWT based filtering method (Shukla et al., 2018b).

2. Hamming window is applied on the filtered EDA signal to enable the analysis over short window durations. Given sampling frequency of f, the recommended values of frame size (N) is, $N = 2 \times f$ and of overlapping window duration (M) is, $M = 0.5 \times f$. A frame size (N) of 2 second along with an overlapping duration (M) of 0.5 sec was selected for the windowing process.

3. For each window, the frequency spectrum was obtained by applying a Fast Fourier Transform (FFT).

4. The frequency spectrum was then mapped onto the Mel-scale through Mel-filters to obtain the Mel-spectrum. The Mel-scale mapping from the actual frequency f can be given as follows:

$$f_{mel} = 2595 \log_{10}(1 + \frac{f}{700}) \tag{8.21}$$

5. Next we take the log of the Mel-spectrum values.

6. Finally, CA was required as per equation 8.20 on the Mel-spectrum to obtain the MFCC features. While applying equation 8.20, the logarithm computation was committed since it was computed in the previous processing step 5. Also, instead DTFT, the discrete cosine transform (DCT) was applied because the absolute value of the Mel-

spectrum is real and symmetric. Hence, for a windowed frame of EDA signal the final MFCC coefficients $C[n]$ are computed by

$$C[n] = \frac{1}{R}\sum_{r=1}^{R} \log(MF_m[r]).\cos\left[\frac{2\pi}{R}\left(\frac{r+1}{2}\right)n\right] \tag{8.22}$$

where R is the number of Mel-filters, $MF_m[r]$ is the Mel-spectrum of the frame r.

For the purposes of MFCC extraction, we selected last 13 components only since the other carry little information. For a given EDA signal, the number of MFCC coefficients obtained via above process resulted in $13 \times Num_{frames}$, where $Num_{frames} = ceil(((length(EDA_{signal}) - N)/M))$.

8.2 Feature Selection

Processing of high dimensional data demands significant computational and space complexity. Therefore, extracting emotion information from high dimensional EDA data can be challenging, especially if the processing is to be done online. Furthermore, many of the EDA features can be irrelevant for the emotion classification or they might be totally redundant. Hence, it is important to automatically identify meaningful smaller subsets of these EDA features to achieve efficient emotion recognition from EDA signals. EDA features have never been explored in detail by previous studies and hence, feature selection techniques were never applied to EDA data.

Feature selection methods can be generally divided into classifier-dependent ('wrapper' and 'embedded' methods), and classifier-independent ('filter' methods) classes (Brown et al., 2012). Wrapper and embedded methods are computationally expensive and both use quite a strict model structure assumptions and hence may produce classifier specific feature subsets. In contrast, filter methods produce generic feature subsets as they are model independent and hence we employed filter methods for selection of significant EDA features.

8.2.1 Joint Mutual Information

Joint Mutual Information (JMI) criterion (Yang and Moody, 1999) provides the best trade-off in terms of accuracy, stability, and flexibility (Brown et al., 2012). JMI focuses on increasing complementary information between features. The JMI score for feature X_k is

$$J_{JMI}(X_k) = \sum_{X_j \in S} I(X_k X_j; Y) \tag{8.23}$$

This is the information between the target Y and a joint random variable X_kX_j, defined by pairing the candidate X_k with each feature X_j previously selected. The candidate feature X_k that maximizes this mutual information is chosen and added to the feature subset S. JMI offers two significant advantages (Yang and Moody, 1999):

1. JMI can distinguish among features even when all of them have same mutual information (MI) and

2. JMI can eliminate the redundancy in the features when one feature is a function of other features.

8.2.2 Conditional Mutual Information Maximization

Conditional Mutual Information Maximization (CMIM) (Fleuret, 2004) is a versatile filter measure and is good measure for general feature selection problems (Freeman et al., 2015). The CMIM criterion for each feature X_k is measured as

$$J_{CMIM}(X_k) = \min_{X_j \in S}[I(X_k : Y|X_j)] \tag{8.24}$$

$I(X_k : Y|X_j)$ is the conditional mutual information between candidate X_k and the target Y given X_j. The candidate feature X_k that minimizes this conditional mutual information is chosen and added to the feature subset S which means that it caries information about the class that is not already captured by the features in the selected set. CMIM can properly identify truly redundant features and noisy features, and gives preference to informative, uncorrelated features (Freeman et al., 2015).

8.2.3 Double Input Symmetrical Relevance

Double Input Symmetrical Relevance (DISR) (Meyer and Bontempi, 2006) is a normalized variant of JMI. DISR takes into consideration the variable complementarity and a lower bound on the mutual information. It uses following modification of the JMI criterion

$$J_{DISR}(X_k) = \sum_{X_j \in S} \frac{I(X_kX_j;Y)}{H(X_kX_jY)} \tag{8.25}$$

DISR criterion motivates the selection of complementary variable of an already selected one with a higher probability. Among several, information theoretic feature selection methods, only JMI, CMIM and DISR satisfies three desirable characteristics (includes reference to a conditional redundancy term, balances the relevance and redundancy terms and uses a low dimensional approximation) of an information based selection criterion (Brown et al., 2012). Hence, we have included DISR for the sake of completeness of our empirical investigation.

8.3 Feature Comparison

8.3.1 AMIGOS Database

The AMIGOS dataset (Miranda-Correa et al., 2017) is a publicly available dataset containing, among other multimodal data, measures of EDA from two experiments: one experiment in which participants watched short ($< 250s$) emotional videos (40 participants), and other experiment in which participants (alone or in groups of four) watched longer ($> 14min$) videos able to elicit diverse emotional states (37 participants, 17 of them in individual setting and 20 in groups). Among other annotations, the dataset includes annotations for emotional arousal and valence of participants in both experiments, provided by three external observers that visually inspect frontal videos of participant's faces during the viewing, and provide an annotation for every 20 seconds segment of the viewing. A total of 12,580 clips were annotated this way (340 clips by 37 participants). The arousal and valence scales used for these annotations were continuous and ranged from -1 (low arousal or valence) to +1 (high arousal or valence), and the agreement between annotators was very good for both the variables (Cronbach's α = 0.96 for arousal, and Cronbach's α = 0.98 for valence).

8.3.2 Classification

EDA is relatively slow response system. The latency of the gradual changes in principle EDA components against an elicited stimulus is between 1.0 and 3.0s (Dawson et al., 2007). Since the focus of our research is to analyze temporal evolution of arousal from EDA signals, we employed external affect annotations of the AMIGOS dataset. We aggregated the ratings from all the three annotators to form a single rating value of more significant meaning for each annotated clip. Since the valence and the arousal scales were continuous and ranged from -1(low arousal) to 1(high arousal), we categorized rating values to LOW or HIGH valence/arousal depending upon if the rating values were less or greater than the mean value of the scale 0.00.

Based upon above categorization, 9886 samples were assigned to class LOW and 2694 to class HIGH for arousal labeling and 9566 were assigned to class LOW and 3014 to class HIGH for valence labeling. This provided an imbalanced two-class dataset. Hence, we used ADASYN method to improve class balance towards equally-sized classes. ADASYN uses a weighted distribution for different minority class samples according to their level of difficulty

Table 8.3 Feature Vector Dimension

Domain	Feature Vector	Number of Features
Time Domain	Event Related	7
	Statistical Features	8
	Hjorth Features	2
	Higher Order Crossings	5
Frequency Domain	Statistical Features	8
	Band Power	9
Time-Frequency Domain	Discrete Wavelet Transform (DWT) Coefficients	56
	Stationary Wavelet Transform (SWT) Features	40
	Mel-frequency cepstral coefficients (MFCC)	481
	Statistical features related to MFCC	5
Total		**621**

in learning and it generates more synthetic data for minority class samples that are harder to learn in comparison to those minority samples that are easier to learn (He et al., 2008). Adaptively generating synthetic data samples in this manner reduces the bias introduced by the imbalanced data distribution.

Table 8.3 describes the feature extraction and feature vector dimension for each subject. A feature matrix was generated from the EDA data of 340 clips for each of the two annotated classes. Features were extracted from all the three domains leading to a total of 660 features. We z-normalized the features to have mean 0 and to have standard deviation 1. The problem of singularities may occur for FS methods (Jenke et al., 2014). Hence, we removed all almost identical features which produced a correlation coefficient higher than .98.

In order to determine the most suitable order for the HOC features, we performed an iterative classification step. We computed the classification rate for several orders of HOC features using stratified 10-fold cross-validation. Classification of the HOC feature data was performed by means of quadratic discriminant analysis (QDA) with diagonal covariance estimates (i.e. Naive Bayes). Figure 8.2 shows the plot of the HOC order vs corresponding classification rate for arousal recognition and for valence recognition. As it is clear from the plot, HOC gets its highest classification rate at order value 5 for both the arousal and the valence recognition, hence we choose the HOC order as 5 for the AMIGOS dataset.

The overall methodology of our recognition system is illustrated in figure 8.3. The sample set was partitioned for each of the 37 participants individually (containing 340 samples for each participant) and collectively for ALL participants using the whole dataset

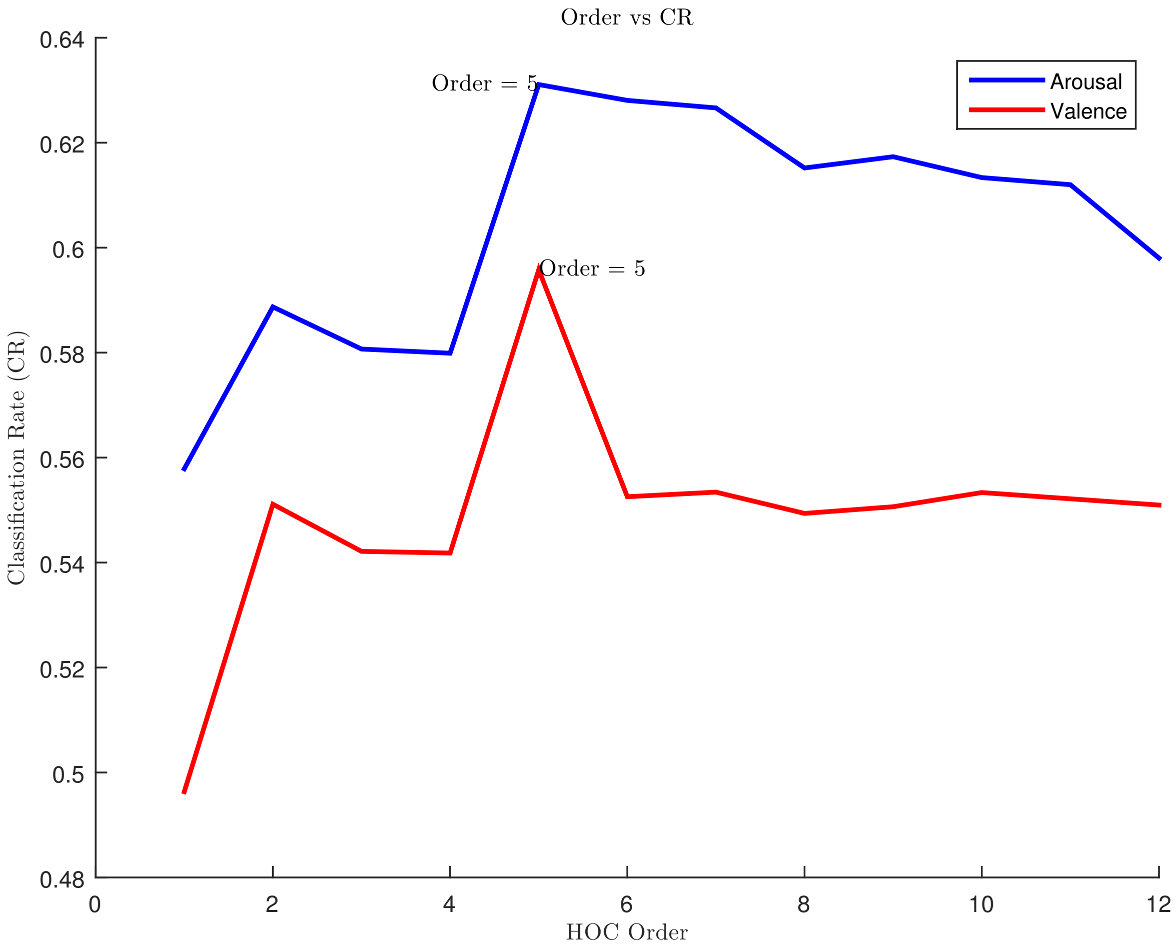


Fig. 8.2 HOC order vs classification rate

(12580 samples). ADASYN method was applied for each partition data to remove the class imbalance and the data then was divided into the ratio of 70 : 15 : 15 for training, validation and testing respectively. We evaluated each of the listed FS methods for each subject individually based on the classification accuracy with a support vector machine (SVM) classifier with radial basis function (RBF) kernel. We employed grid search and 3 fold cross validation method to determine the optimal regularization parameter C and also to determine the free parameter γ of the Gaussian RBF. We applied SVM since SVM has been reported to provide the best classification accuracy during the recognition of affective states from physiological cues (Hosseini et al., 2010; Li and Lu, 2009; Rani et al., 2006).

8.4 Results

We aim to provide insights on the following questions:

1. How the different feature selection algorithms perform and what is the optimal number of features?

2. Which features are selected most frequently for the arousal recognition?

3. Which features are selected most frequently for the valence recognition?

4. How the performance of the subject-depedent classification differs from the subject-independent classification?

8.4.1 Feature Selection Methods

Table 8.4 presents the optimal accuracies for arousal recognition, F1 score and the optimal number of features in tabular form across all the 37 subjects and the three FS methods. The optimal accuracies are the highest accuracies obtained at the optimal number of features. Table 8.5 shows the similar results for the valence recognition. Table 8.4 and table 8.5 also

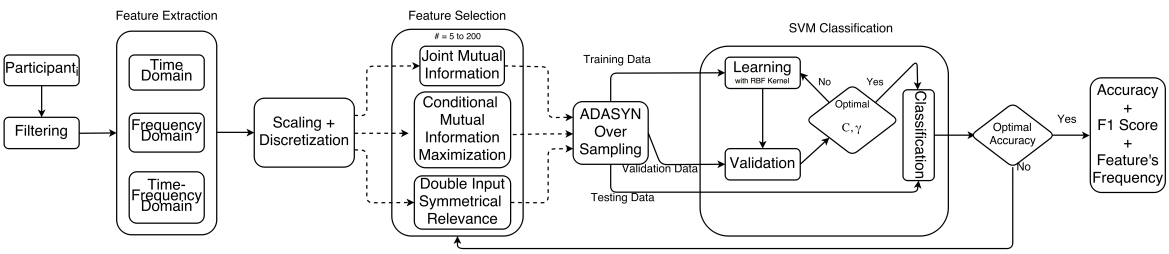


Fig. 8.3 Block diagram of supervised classification system

present the average results for all 37 subjects across all three FS methods. In addition,we have included subject-independent (ALL) classification results also for arousal and valence recognition in Table 8.4 and table 8.5 respectively.

Statistical Analysis

We tested if the values of F1-scores provided by the different selection algorithms are significantly higher than 0.5 (at the $p < .05$ level). In the case of arousal detection, the JMI ($M = 0.58; SD = 0.17; t(36) = 2.78; p = .009$), the CMIM ($M = 0.63; SD = 0.1; t(36) = 7.92; p < .001$), and the DISR ($M = 0.64; SD = 0.11; t(36) = 7.49; p =< .001$) algorithms provided mean values of F1-scores that were significantly higher than 0.5. Also in the case of valence, the JMI ($M = 0.61; SD = 0.1; t(36) = 6.49; p < .001$), CMIM ($M = 0.59; SD = 0.13; t(36) = 4.32; p < .001$), and DISR ($M = 0.63; SD = 0.09; t(36) = 8.85; p < .001$) algorithms provided values significantly higher than 0.5 for F1-scores.

We also compared the results of subject-independent (ALL) classification with subject-dependent (37 individual subjects) across all three FS methods for arousal recognition. The results of the subject-dependent recognition results were significantly higher than subject-independent (ALL) results in terms of accuracy (JMI: $t(36) = 17.1888, p < 2.2e-16$, CMIM: $t(36) = 18.4018, p < 2.2e-16$, DISR: $t(36) = 20.5805, p < 2.2e-16$) and in terms of average F1-scores (JMI: $t(36) = -1.5847, p = 0.9391$, CMIM: $t(36) = -1.1653, p = 0.8742$, DISR: $t(36) = 0.8217, p = 0.2083$). In terms of optimal number of features, the results of the subject-dependent recognition results were significantly lower than subject-independent (ALL) results (JMI: $t(36) = -8.184, p = 4.912e-10$, CMIM: $t(36) = -5.0724, p = 6.019e-06$, DISR: $t(36) = -11.0016, p = 2.283e-13$).

A similar pattern was observed on comparing the results of subject-independent (ALL) classification with subject-dependent (37 individual subjects) for the valence recognition. The results of the subject-dependent recognition results were significantly higher than subject-independent (ALL) results in terms of accuracy (JMI: $t(36) = 17.389, p < 2.2e-16$, CMIM: $t(36) = 18.3873, p < 2.2e-16$, DISR: $t(36) = 21.8212, p < 2.2e-16$) and in terms of average F1-scores (JMI: $t(36) = -0.8515, p = 0.7999$, CMIM: $t(36) = -1.7024, p = 0.9513$, DISR: $t(36) = 3.5855, p = 0.0004953$). In terms of optimal number of features, the results of the subject-dependent recognition results were significantly lower than subject-independent (ALL) results (JMI: $t(36) = -10.3046, p = 1.385e-12$, CMIM: $t(36) = -4.7272, p = 1.724e-05$, DISR: $t(36) = -6.31, p = 1.341e-07$).

Finally, the outputs of the three feature selection algorithms in terms of accuracy, F1-score, and optimal number of features, were compared using a series of repeated measures (within-subjects) ANOVAs, in order to examine if there is a significant difference between them.

Table 8.4 Accuracy, F1 Score, Optimal Number of Features for Different FS Methods Listed for Each Subject for Arousal Recognition

Participant	JMI			CMIM			DISR			Average		
	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+
1	77.5	0.66	37	76.5	0.68	145	76.5	0.66	63	76.8	0.7	81.7
2	80.4	0.49	144	81.4	0.5	165	83.3	0.61	127	81.7	0.53	145.3
3	88.2	0.54	13	92.2	0.58	26	88.2	0.54	153	89.5	0.53	64
4	82.4	0.6	77	80.4	0.63	81	80.4	0.53	70	81.1	0.57	76
5	85.3	0.74	159	86.3	0.71	171	86.3	0.79	60	86	0.73	130
6	88.2	0.72	34	88.2	0.75	136	86.3	0.53	48	87.6	0.67	72.7
7	73.5	0.63	92	79.4	0.62	60	77.5	0.64	6	76.8	0.6	52.7
8	86.3	0.57	137	89.2	0.55	67	88.2	0.54	105	87.9	0.57	103
9	93.1	0.6	36	93.1	0.59	27	93.1	0.59	46	93.1	0.6	36.3
10	90.2	0.56	131	91.2	0.57	16	92.2	0.58	13	91.2	0.6	53.3
11	82.4	0.54	88	81.4	0.62	191	84.3	0.65	107	82.7	0.6	128.7
12	80.4	0.49	183	81.4	0.53	185	82.4	0.58	25	81.4	0.53	131
13	75.5	0.55	53	73.5	0.59	180	74.5	0.54	167	74.5	0.57	133.3
14	93.1	0.59	130	94.1	0.73	106	91.2	0.57	61	92.8	0.63	99
15	83.3	0.58	142	87.3	0.62	122	84.3	0.51	155	85	0.57	139.7
16	87.3	0.62	197	87.3	0.53	167	87.3	0.62	39	87.3	0.57	134.3
17	83.3	0.64	62	83.3	0.66	162	85.3	0.7	14	84	0.67	79.3
18	78.4	0.52	200	78.4	0.7	194	80.4	0.7	13	79.1	0.63	135.7
19	97.1	0.69	86	98	0.75	6	96.1	0.66	123	97.1	0.73	71.7
20	90.2	0.66	92	91.2	0.63	174	93.1	0.84	105	91.5	0.7	123.7
21	82.4	0.74	70	85.3	0.74	83	85.3	0.76	42	84.3	0.73	65
22	100	1	186	100	1	166	100	1	195	100	1	182.3
23	85.3	0.52	131	82.4	0.5	14	82.4	0.6	161	83.4	0.53	102
24	93.1	0.66	145	93.1	0.59	65	94.1	0.68	15	93.4	0.67	75
25	87.3	0.53	6	88.2	0.54	24	89.2	0.55	138	88.2	0.53	56
26	88.2	0.54	58	93.1	0.59	92	88.2	0.54	28	89.8	0.53	59.3
27	84.3	0.51	197	88.2	0.63	69	85.3	0.56	87	85.9	0.57	117.7
28	86.3	0.57	158	84.3	0.59	190	88.2	0.67	39	86.3	0.63	129
29	78.4	0.71	183	80.4	0.71	187	81.4	0.74	139	80.1	0.7	169.7
30	77.5	0.56	75	77.5	0.64	31	76.5	0.65	155	77.2	0.63	87
31	72.6	0.59	7	76.5	0.63	7	78.4	0.72	10	75.8	0.63	8
32	89.2	0.6	53	89.2	0.55	197	91.2	0.63	20	89.9	0.6	90
33	70.6	0.69	26	68.6	0.67	53	68.6	0.69	38	69.3	0.7	39
34	83.3	0.51	185	84.3	0.51	48	82.4	0.5	101	83.3	0.5	111.3
35	92.2	0.58	162	91.2	0.57	87	96.1	0.66	24	93.2	0.63	91
36	89.2	0.68	36	89.2	0.55	77	87.3	0.53	74	88.6	0.6	62.3
37	97.5	0.7	81	96.1	0.66	71	98	0.75	90	97.2	0.73	80.7
Average	85.2	0.61	104	86	0.63	104	86	0.64	77	**85.75**	**0.63**	**95.04**
ALL	65.1	0.65	187	64.6	0.65	158	62.6	0.62	176	**64.1**	**0.64**	**173.6**

A*– Classification Accuracy F1#– F1 Score F+– Optimal Number of Features

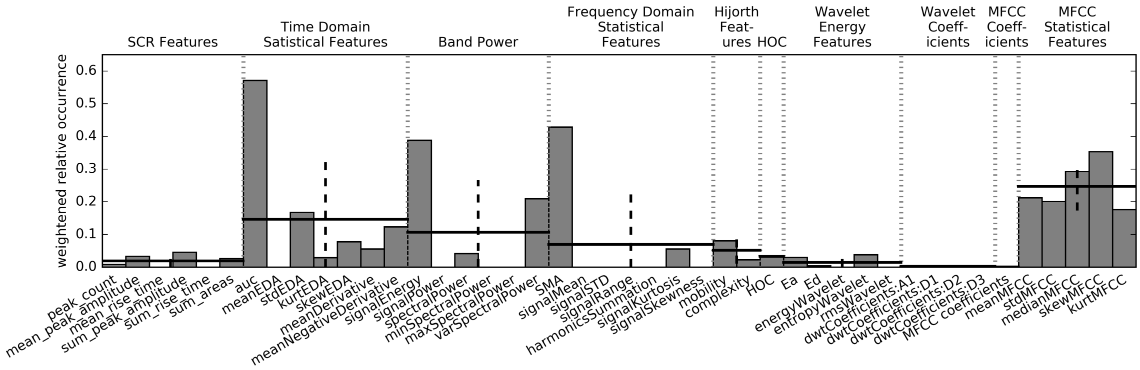


Fig. 8.4 Weighted relative frequency of each EDA feature type for arousal recognition. Black horizontal bar represents the mean for the group, and the vertical dashed black line represents the standard deviation for each group.

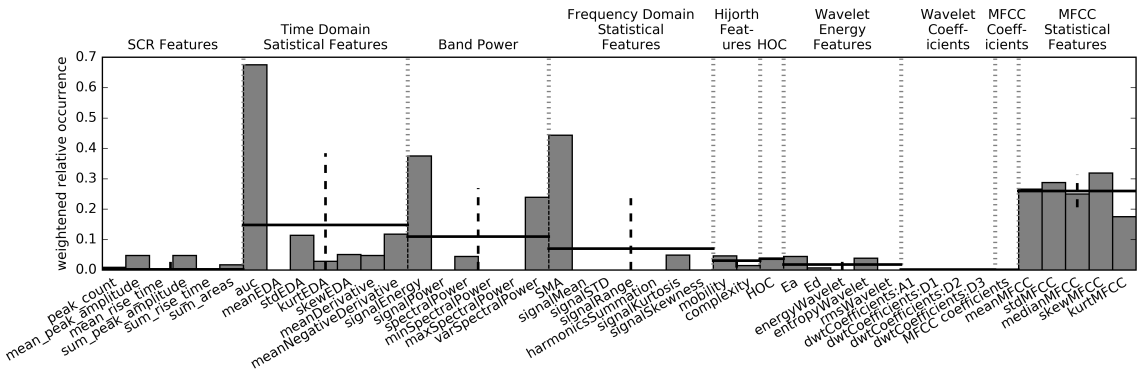


Fig. 8.5 Weighted relative frequency of each EDA feature type for valence recognition. Black horizontal bar represents the mean for the group, and the vertical dashed black line represents the standard deviation for each group.

Regarding arousal, no significant difference (at $p < .05$ level) was found between the three algorithms in regarding their accuracy ($F(2,105) = 0.1; p = .91$), F1-scores ($F(2,105) = 0.29; p = .75$), or optimal number of features ($F(2,105) = 1.07; p = .35$). Neither in the case of valence there were significant differences in terms of the accuracy ($F(2,105) = 0.01; p = .99$), F1-score ($F(2,105) = 0.17; p = .85$), or optimal number of features ($F(2,105) = 0.39; p = .68$). That is, neither of the three outputs is significantly different between the three algorithms, or, in other words, there is no evidence that one of the algorithms outperforms the others.

Table 8.5 Accuracy, F1 Score, Optimal Number of Features for Different FS Methods Listed for Each Subject for Valence Recognition

Participant	JMI			CMIM			DISR			Average		
	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+	A*	F1#	F+
1	70.6	0.58	20	68.6	0.55	171	68.6	0.61	164	69.3	0.58	118.3
2	76.5	0.58	68	77.5	0.54	73	77.5	0.58	192	77.1	0.57	111
3	87.3	0.53	122	85.3	0.52	7	82.4	0.54	140	85	0.53	89.7
4	75.5	0.6	155	76.5	0.63	65	78.4	0.59	41	76.8	0.61	87
5	91.2	0.78	75	88.2	0.74	104	89.2	0.79	152	89.5	0.77	110.3
6	85.3	0.56	45	87.3	0.53	50	84.3	0.59	116	85.6	0.56	70.3
7	66.7	0.67	19	66.7	0.65	20	67.6	0.67	93	67	0.66	44
8	74.5	0.49	195	82.4	0.54	132	81.4	0.57	13	79.4	0.53	113.3
9	85.3	0.52	132	91.2	0.57	124	94.1	0.68	12	90.2	0.59	89.3
10	84.3	0.56	139	86.3	0.53	53	86.3	0.61	105	85.6	0.56	99
11	75.5	0.63	92	77.5	0.65	90	77.5	0.64	190	76.8	0.64	124
12	76.5	0.5	5	73.5	0.62	113	72.5	0.53	69	74.2	0.55	62.3
13	69.6	0.57	53	75.5	0.71	49	76.5	0.67	161	73.9	0.65	87.7
14	95.1	0.63	44	91.2	0.57	185	91.2	0.57	58	92.5	0.59	95.7
15	81.4	0.5	112	83.3	0.55	76	81.4	0.59	82	82	0.55	90
16	77.5	0.54	156	81.4	0.5	72	80.4	0.56	47	79.7	0.53	91.7
17	77.5	0.6	126	78.4	0.54	111	80.4	0.68	46	78.8	0.61	94.3
18	88.2	0.74	59	88.2	0.67	39	89.2	0.6	55	88.6	0.67	51
19	91.2	0.57	147	92.2	0.69	165	93.1	0.59	109	92.2	0.62	140.3
20	93.1	0.84	94	92.2	0.8	199	94.1	0.85	160	93.1	0.83	151
21	77.5	0.66	93	80.4	0.72	162	79.4	0.71	72	79.1	0.7	109
22	86.3	0.53	12	85.3	0.52	89	84.3	0.51	167	88.6	0.52	89.3
23	75.5	0.5	64	78.4	0.59	16	79.4	0.55	7	77.8	0.55	29
24	95.1	0.63	61	93.1	0.71	24	93.1	0.71	16	93.8	0.69	33.7
25	77.5	0.56	56	79.4	0.52	123	81.4	0.59	45	79.4	0.56	74.7
26	87.3	0.53	188	86.3	0.53	79	85.3	0.52	161	86.3	0.53	142.7
27	88.2	0.59	51	91.2	0.57	138	91.2	0.63	29	90.2	0.6	72.7
28	85.3	0.6	71	83.3	0.55	178	83.3	0.58	86	84	0.58	111.7
29	92.2	0.85	171	90.2	0.72	106	91.2	0.82	198	91.2	0.8	158.3
30	79.4	0.58	30	80.4	0.56	134	78.4	0.63	75	79.4	0.59	79.7
31	87.3	0.8	7	78.4	0.59	156	83.3	0.76	18	83	0.72	60.3
32	90.2	0.56	167	92.2	0.73	61	97.1	0.83	142	93.1	0.7	123.3
33	84.3	0.56	183	83.3	0.55	140	85.3	0.52	160	84.3	0.54	161
34	83.3	0.58	31	84.3	0.56	184	87.3	0.53	123	85	0.56	112.7
35	88.2	0.54	133	91.2	0.57	36	94.1	0.61	51	91.2	0.57	73.3
36	93.1	0.8	6	92.2	0.58	142	93.1	0.71	94	92.8	0.7	80.7
37	86.3	0.53	166	94.1	0.68	51	93.1	0.66	170	91.2	0.62	129
Average	83.2	0.6	90.5	84.0	0.61	100.0	84.5	0.63	97.8	**83.9**	**0.61**	**96.3**
ALL	61.9	0.62	190	62.8	0.63	142	58.0	0.58	159	**60.9**	**0.61**	**163.6**

A^*– Classification Accuracy $F1^\#$– F1 Score F^+– Optimal Number of Features

Table 8.6 Comparison of F1-scores

Method	**Proposed**	**(Miranda-Correa et al., 2017)** (EDA)	**(Miranda-Correa et al., 2017)** (EEG)
Arousal	**0.628**	0.548	0.592
Valence	**0.614**	0.531	0.576

8.4.2 Significant Features

The significance of the employed EDA features from all the three domains were tested to identify the most frequently selected features by the different FS methods and to identify the features with the most significant performance for the arousal and the valence recognition. The approach for testing the superiority of the features was similar to the one described in (Jenke et al., 2014). It involves the computation of the relative frequency of each of the feature types, which is obtained via the following process:

1. Firstly we created a histogram of the feature occurrence given the features selected for the optimal number of features across all the subject and the FS methods.

2. Then to take into account the random assignment of the features, each bin of the histogram was normalized by dividing the occurrence of each feature type by the feature's cardinality (e.g. 5 in the case of HOC).

3. Then while averaging, these relative frequencies were weighted across the FS methods by multiplying with the achieved classification accuracy.

Employing above statistic, the relative frequency of the feature types vary between 0 and 1 and the most significant features score higher than the non-significant features for emotion recognition. Figure 8.4 shows the weighted relative frequency of each feature type for arousal recognition. Figure 8.5 shows the weighted relative frequency of each feature type for valence recognition. The results are analyzed separately for arousal and valence recognition patterns.

The most frequently selected feature group for arousal recognition is the MFCC Statistical features. Statistical features related to SCR of the EDA signal in the time domain and the band power related features in the frequency domain performed second best among all feature groups. The single most performing feature among all the employed features is the AUC feature followed by SMA and the signal energy feature. The standard deviation or variance of the SCR signal along with the derivative features of the SCR signal also yield a better performance over other time domain statistical features. In general, the statistical features of the SCR signal perform better than the SCR related features of the EDA signals in the time domain. It is worth noting that the time domain statistical features, band power and the frequency domain statistical features show a higher variance over the other feature groups. It indicates that particular feature types in these groups are more valuable than others. These feature types are AUC, signal energy and SMA in time domain statistical features, band power and the frequency domain statistical features, respectively. The variance among the MFCC statistical features are lower in comparison to the 3 other feature types. It means that all the features in this feature type are significant. The least frequently selected features are the coefficients of the wavelet and the MFCC as they have low weighted relative occurrence scores.

On comparison, we found that the feature usage for valence recognition follows the same trend as that of the arousal recognition. The most frequently selected feature group for valence recognition is the MFCC Statistical features which is same as the arousal recognition. It is again followed by the time domain and the band power related features in the frequency domain. The single most performing feature among all the employed features is also the AUC feature followed by SMA and the signal energy feature.

8.4.3 Validation

To assess the validity of the obtained results, we compared the classification performance of our results with that of researchers of AMIGOS dataset (Miranda-Correa et al., 2017). We have used only F1-scores for the comparison as the researchers from (Miranda-Correa et al., 2017) have not provided the classification accuracies. The average F1-scores of our results for all classes over all participants are given in the second last row of table 8.4 and table 8.5 and are reproduced in table 8.6 along with the F1-scores as provided by (Miranda-Correa et al., 2017) for corresponding short videos classification. Please note that F1-scores in table 8.4 and table 8.5 are given with two decimal places to increase the readability. But, we have provided the average F1-score with three decimal places in table 8.6 for an accurate comparison. We performed an independent one-sample $t-test$ to see if the F1-scores of our classification results is significantly higher ($p < .05$) than results of (Miranda-Correa et al., 2017). In case of both arousal ($t(36) = 4.4678; p =< .001$) and valence recognition ($t(36) = 6.4671; p =< .001$), F1-scores of our classification is significantly higher than (Miranda-Correa et al., 2017). The researchers have also reported the highest recognition performance using EEG signal modalities on short videos of AMIGOS dataset(Miranda-Correa et al., 2017). Hence, we included recognition results of EEG modalities also in table 8.6 for comparison. Applying the independent one-sample $t-test$, we found that F1-

scores of our recognition results are significantly higher than that of EEG modalities, for both arousal ($t(36) = 1.8511; p =< .05$) and valence recognition ($t(36) = 2.9684; p =< .01$).

8.5 Discussion

While no difference in the performance of any single feature selection method over the others was found, all the three different feature selection methods yielded a higher average classification accuracy and F1 scores for both the arousal and as well as the valence recognition. It indicates the significant potential of selectively chosen EDA features to be able to actively detect the patterns of arousal and valence recognition.

Superior subject-dependent classification accuracies with better average F1-scores using a lower number of features were obtained in comparison to the subject-independent classification. The big difference in such classification results is not surprising. It is known that subjects with different psychophysiological profiles tend to have different physiological responses to the same stimuli Henriques and Paiva (2014). In addition, non-emotional individual contexts vary in a complex manner among different subjects Kim and André (2008). If the subject is known in advance to the system or if the system can undergo a learning phase for each subject prior to the classification, then the emotion classification can be done in a user-dependent way. We believe that this is one of the biggest challenges of real-time emotion recognition and it goes beyond the scope of this article.

SCR features have been widely used in the literature but it yielded low wighted relative occurrence scores in our study. Our study revealed that among all SCR features amplitude of the SCR peaks is the most significant feature for the recognition of both the arousal and the valence from EDA signals. We also demonstrated that the commonly used rise time feature does not play an important role for the same. While AUC of the EDA signal has not been highly exploited in the literature, yet as a single feature, it yields the highest performance for emotion recognition from EDA signals. It is also true for SMA and signal energy features of the EDA signal. The most significant findings of our study are the performance of the statistical features related to MFCC. It outperformed all other feature types across all the three domains.

In comparison to the classification results presented in (Miranda-Correa et al., 2017), we obtained significantly better results over the same dataset and even against other modalities. This indicates a higher confidence in results presented by us and confirms that the selection of appropriate features plays a significant role in the emotion recognition using EDA signals.

8.6 Conclusion

In this chapter, we reviewed different EDA features for emotion recognition suggested by 25 studies. We systemically analyzed the significance and the suitability of different EDA feature across time, frequency and time-frequency domains using three feature selection methods, JMI, CMIM and DISR methods and employing machine learning techniques on a publicly available AMIGOS dataset.

We did not find any significant difference in the performance of the three feature selection methods employed in this study. All the three methods indicated use of $\sim$95 features on an average for arousal recognition and of $\sim$96 features on an average for valence recognition. The results reported an average accuracy of 85.75%(F1-score: 0.63) for arousal recognition and an average accuracy of 83.9%(F1-score: 0.61) for valence recognition. Also, the subject-dependent classification results were significantly higher than the subject-independent classification for both the arousal and the valence recognition. Statistical MFCC features along with the AUC and SMA features outperformed the commonly used SCR and statistical features related to SCR of the EDA signal in time domain. The classification results were also found to be superior to the classification results presented in (Miranda-Correa et al., 2017) using EDA signal. Our results also outperformed the best classification results presented in (Miranda-Correa et al., 2017) using EEG signal modalities.

Subjects with different psychophysiological profiles tend to have different physiological responses for the same stimuli (Henriques and Paiva, 2014). Failure to address this individual variability can negatively affect the classification performance of emotional state which is visible by the results of subject-independent classification in our results. To overcome this issue, a general model for the emotion classification can be prepared with sufficiently high number of subjects and such a model can then be fine tuned with the baseline values of the new user. In addition, online emotion recognition demands temporal evolution analysis of EDA signals which is facilitated by annotations over short durations. Hence, more baseline datasets with short duration annotations are required to enable to community to advance online emotion recognition and to address individual variability.

Chapter 9

Conclusions

This chapter first describes the principal contributions of this research, followed by several lessons learned by the author, limitation of the present research and several important directions for the future work. A brief summary of the challenges tackled together with the most important results is given at the end of the chapter.

9.1 Contributions

This dissertation addresses the problem of designing and evaluating the effectiveness of robot-assisted interventions for cognitive stimulation among individuals with ID and the development of automated and online emotion recognition from physiological signals. This is a challenging endeavor because of the challenges arising from both an engineering and cognitive rehabilitation point of view in this interdisciplinary field of research. In the light of these challenges, the dissertation makes following five principal contributions:

1. **Design of robot-assisted cognitive stimulation.** Based on the interviews with the experts, I have designed a general framework incorporating the robot-assisted therapy for cognitive rehabilitation. This framework is based on the participatory design approach and hence addresses the user's cognitive stimulation needs. I have also identified and listed the feasibility of robot interventions, possible usage scenarios and the metrics that can be used for evaluation of such robot interactions in stimulation context. The proposed use-case and the evaluation metrics take into account the requirements and benefits for both the user and their caregiver and hence can save the time in the implementation of the robot-assisted cognitive stimulation activities and help the caregiver accomplish his/her task of stimulating the user more effectively.

2. **Impact evaluation of robot-assisted cognitive stimulation on the users and their caregiver.** To prove the efficacy of the robot empowered cognitive rehabilitation, I evaluated the impact of robot-assisted interventions on both the users and their caregivers. I have demonstrated the positive effects of such interventions on the users, mainly an increase in the engagement. I argued about the utility of robots over monitors and have indicated that robot embodiment is more beneficial to simple tablets or screens. I have also displayed the significant reduction in caregivers burden during such interventions. These evaluations have provided valuable insights to take maximum advantage of SAR in health care.

3. **Multi-modal database for emotion recognition.** To fuel the research on the emotional adaptive behavior of robots for the cognitive stimulation of users with ID, I have presented MuDERI which is a first-ever annotated multimodal database of individuals with ID. MuDERI is an annotated multimodal database of audiovisual recordings, RGB-D videos and physiological signals of 12 participants in actual settings, which were recorded as participants were elicited using personalized real-world objects and/or activities. To benefit other researchers in the field from using the database, I have made MuDERI publicly available by request. MuDERI will help to facilitate develop and test new algorithms for real-time emotion recognition among individuals with ID.

4. **Real-time denoising of EDA signals.** To denoise the EDA signal during online processing, I have proposed an efficient wavelet-based method for artifacts attenuation while minimizing distortions. I tested the proposed method on EDA recordings from publicly available driver dataset collected during real-world driving, on EDA recordings from MuDERI dataset and the results were compared to those of three state-of-the-art methods for EDA signal filtering. In addition, I tested the proposed method for the online filtering of EDA signals collected while twelve volunteers conducted tasks designed to elicit various stress states. The results demonstrated that the proposed algorithm outperforms existing approaches and it has a lower computational cost. The lower computational requirements of the proposed method along with no external dependencies can facilitate the embedding of denoising algorithms into wearable devices. It can, therefore, boost the creation of applications able to monitor user's cognitive and emotional states.

5. **EDA feature extraction and selection for emotion recognition.** To identify the significant EDA features, I have reviewed feature extraction methods for emotion recognition from EDA based on 25 studies. I compared these features for feature selection using machine learning techniques on a publicly available AMIGOS dataset.

I presented the results of the performance of three feature selection methods and usage of selected feature types across time, frequency and time-frequency domains. The results did not find any statistical evidence was found that one of the feature selection methods outperforms the others. However, the subject-dependent classification results were significantly higher than the subject-independent classification for both the arousal and the valence recognition. MFCC and related statistical features were explored for the first time for the emotion recognition from EDA signals and they outperformed all other feature types, including the most commonly used SCR related features. I also compared the results with methods employed by researchers of AMIGOS dataset for classification of emotional states and they show that the EDA features explored in this study provided better performance.

9.2 Lessons Learned

The author made several mistakes during the completion of the thesis and learned several lessons which are listed below:

- **Minimize external influence.** Due to their mental conditions, individuals with ID can get distracted easily. Hence, while designing trials involving them, minimizing external influences is highly recommended. This can include a noise-free environment, stable and comfortable temperature and light conditions, limiting the total experiment duration to $10-15$ minutes. Along the same lines, the preprocessing steps for a trial involving robot interaction among individuals with ID, such as making the robot ready, charging sensors, preparing camera etc., must be performed before the arrival of the participants to reduce the nervousness among the participants.
- **Robot movements can be troublesome.** Movement of the robots during the interaction with individuals with mental health concerns can be troublesome since robot movements might not be reliable. Unreliable robot movements can cause harm to the users (and to themselves!) and hence these movements must be planned very carefully around individuals with ID. Robots with wheeled locomotion are recommended over bipedal robots.
- **Provide thorough instructions to the caregivers.** Caregivers play an important role in the trials involving individuals with mental health concerns. Since the research on SAR empowered cognitive stimulation is still in its infancy and it might involve technical details hence the caregivers need to be provided with thorough instructions

about the trial execution and their role. It can reduce the errors in the trial and at the same time can make the job easier for the caregiver.

- **Ethical Implications goes beyond physical level.** The human-robot interaction (HRI) of the proposed system works not only at the physical level but also at the cognitive level of the target users. This increases the complexity of the ethical, legal and social (ELS) implications involved. These ELS issues cannot be comprehended with the current regulations and guidelines of the robot industries. Accordingly, a new set of guidelines concerning the ELS issues are required for the cognitive interactions of SAR in therapeutic contexts. Meanwhile, the advisory committee of the institution along with the stakeholders (e.g. guardians, doctors, caregivers etc.) must be involved to identify the ELS issues concerning the trials and the they must be designed and executed accordingly.

9.3 Future Directions

The ultimate objective of this research is to empower the robots with an automated emotion-recognition system allowing emotion adaptive human-robot interaction for cognitive stimulation. Based on the topics considered in this thesis, this objective inspires following interesting areas of future work:

- **The effectiveness of Adaptive Robot Behavior.** Improvement in the effectiveness of robot assisted interventions produced by using emotional adaptive robot behavior needs to be demonstrated by conducting an evaluation study. A specific cognitive stimulation therapy can be selected and literature can be explored to select a set of indicators that have a direct correlation with the effectiveness in the delivery of the selected therapy. Two versions of the stimulation activity can be implemented, with and without the use of emotional adaptive robot behavior. Effectiveness of emotional adaptive robot assisted interventions can be evaluated by comparing the indicators in both the types of the implementations. Such evaluation results can be helpful to achieve a market ready product in a near future.

- **Multi-Modal Fusion.** Fusion of emotion markers from multiple sources such as combining information from EDA, HR and EEG etc. can compensate for weaknesses of one source and can provide better estimation of the emotional state. EDA is usually considered as a correlate of emotional arousal. On the other hand, EEG is widely used in human-computer interaction research for obtaining information on the emotional states, including arousal, valence (Schmidt and Trainor, 2001) and engagement

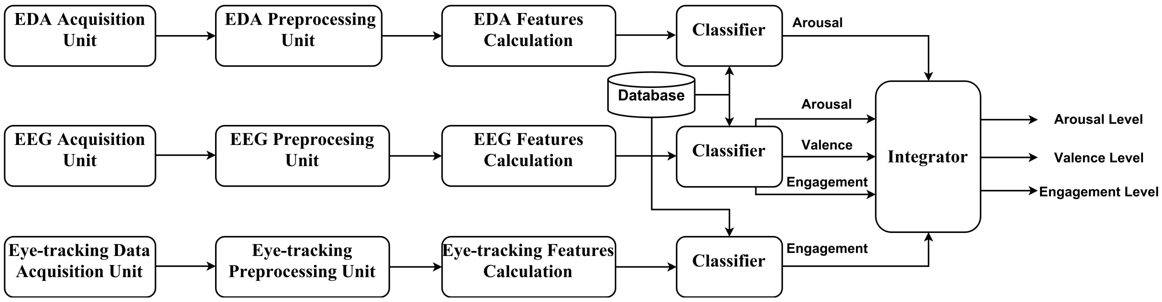


Fig. 9.1 Comphrensive Multimodal Emotion Recognition.

(McMahan et al., 2015). Moreover, the emotional arousal and the valence level may be further combined with external information of the subject, e.g. eye tracking, acquired via an eye-tracking device configured to perform such action. Therefore, the emotional state in terms of human emotion can be represented using a comprehensive model comprising arousal, valence and level of engagement. Fig. 9.1 illustrates this particular embodiment.

- **Address individual variability.** Subjects with different psychophysiological profiles tend to have different physiological responses for the same stimuli (Henriques and Paiva, 2014). Failure to address this individual variability can negatively affect the training, testing and classification performance of emotional state. While it is easy to treat the individual variability for emotion recognition on known users but the real challenge lies to generalize the emotion recognition to unknown users. In such cases, a baseline of the users physiological response can be used to tune the individual user profile.

9.4 Summary

Robot-assisted systems for cognitive rehabilitation can increase the reach of potential benefits of evidence-based psychological or psychosocial interventions to the individuals with a wide range of mental health concerns. Existing researches in SAR lacks clinical validation and hence the medical practitioners have little motivation for their use in clinical practices. Besides existing human-robot interactions are inattentive to the user's current emotional state and engagement. Cognitive rehabilitation interventions for individuals with mental health concerns demand complex human robot interaction and ubiquity of wearable devices motivates for robot interaction systems which can autonomously acquire information about the user's emotional state, intentions and surrounding context so the robot can adapt its interactions accordingly.

In this thesis, I have described the design, implementation of robot-assisted cognitive rehabilitation activities and real-time emotion recognition from EDA signals. Design of robot-assisted interventions presents a coherent framework to produce positive effects on both the users and the caregivers. The implementation of the system confirms an increased engagement among users and a significant reduction in caregivers burden. The development of the emotion recognition algorithms has shown that it is possible to process the EDA signals in real time with minimal lag to infer the emotional state of individuals with ID.

There are various research opportunities that can be exploited. A clinical validation of robot-assisted interventions among individuals with intellectual disability can be performed. The benefits of evidence-based robot-assisted interventions can be extended to individuals with a wide range of mental health concerns, including kids with ASD, elderly people with dementia. The emotion recognition system can be fused with information from other physiological signals and/or speech, video data. The emotion recognition system can be generalized to address individual variability and application of EDA signal based emotion recognition can be conceptualized in several related domains, such as newly emerging field of neuro-marketing.

Addressing the challenges presented in this dissertation will make the benefits of robot-assisted interventions readily available to wider range of user population. The real-time emotion recognition ability will not only provide intriguing possibilities to improve quality of life in health and social care but also in applications related to assistive technologies, learning tools. It can certainly raise the complexity of currently limited human-machine interaction (HMI) to include emotional adaptation that is now one of the key attributes of human interaction only.

The research presented in this dissertation plays an important role towards accomplishing this vision. By developing robot-assisted cognitive rehabilitation interaction and including real-time emotion recognition, I have expanded the scope of SAR in mental health scenarios in which adaptive robot interaction can be integrated. I have further motivated future research that can be benefited by the rich modality of the human emotions expressed in the pure unaltered form of physiological signals, to the advancement of socially and emotionally intelligent interactions.

References

Alberdi, A., Aztiria, A., and Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, 59:49 – 75.

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Publishing, Arlington, VA, 5 edition.

Banks, M. R., Willoughby, L. M., and Banks, W. A. (2008). Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. *Journal of the American Medical Directors Association*, 9(3):173–177.

Barreto, A., Zhai, J., and Adjouadi, M. (2007). Non-intrusive physiological monitoring for automated stress detection in human-computer interaction. In Lew, M., Sebe, N., Huang, T. S., and Bakker, E. M., editors, *Human–Computer Interaction*, pages 29–38, Berlin, Heidelberg. Springer Berlin Heidelberg.

Bellamy, G., Croot, L., Bush, A., Berry, H., and Smith, A. (2010). A study to define: profound and multiple learning disabilities (PMLD). *Journal of intellectual disabilities: JOID*, 14(3):221–235.

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

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

Benesty, J., Sondhi, M., and Huang, Y. (2007). *Springer Handbook of Speech Processing*. Springer Handbook of Speech Processing. Springer Berlin Heidelberg.

Bienstein, C. and Fröhlich, A. (2003). *Basale Stimulation in der Pflege: die Grundlagen*. Edition Pflege. Kallmeyer.

Boucsein, W. (2012). *Electrodermal Activity*. Springer US, Boston, MA.

Bowling, A. (2014). *Research Methods In Health: Investigating Health And Health Services*. Open University Press, Maidenhead New York, NY, 4 ed edition.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimzation Problems*, pages 127–214. Cambridge University Press.

Brown, G., Pocock, A., Zhao, M.-J., and Luján, M. (2012). Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *J. Mach. Learn. Res.*, 13(1):27–66.

Brown, J. F., Brown, M. Z., and Dibiasio, P. (2013). Treating Individuals With Intellectual Disabilities and Challenging Behaviors With Adapted Dialectical Behavior Therapy. *Journal of Mental Health Research in Intellectual Disabilities*, 6(4):280–303.

Cabibihan, J.-J., Javed, H., Ang, M., and Aljunied, S. M. (2013). Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics*, 5(4):593–618.

Carulla, L. S., Reed, G. M., Vaez-Azizi, L. M., et al. (2011). Intellectual developmental disorders: towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11. *World Psychiatry*, 10(3):175–180.

Chellali, R. and Hennig, S. (2013). Is it time to rethink motion artifacts? temporal relationships between electrodermal activity and body movements in real-life conditions. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 330–335.

Chen, W., Jaques, N., Taylor, S., Sano, A., Fedor, S., and Picard, R. (2015). Wavelet-based motion artifact removal for electrodermal activity. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015(5):6223–6.

Chen, Z., Haykin, S., Eggermont, J. J., and Becker, S. (2007). *Appendix E: Expectation–Maximization Algorithm*, pages 384–386. John Wiley Sons, Inc.

Coeckelbergh, M., Pop, C., Simut, R., Peca, A., Pintea, S., David, D., and Vanderborght, B. (2016). A Survey of Expectations About the Role of Robots in Robot-Assisted Therapy for Children with ASD: Ethical Acceptability, Trust, Sociability, Appearance, and Attachment. *Science and Engineering Ethics*, 22(1):47–65.

Cohen, J. (1988). In Cohen, J., editor, *Statistical Power Analysis for the Behavioral Sciences (Revised Edition)*. Academic Press, revised edition edition.

Coifman, R. R. and Donoho, D. L. (1995). *Translation-Invariant De-Noising*, pages 125–150. Springer New York, New York, NY.

Davis, S. and Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4):357–366.

Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). The electrodermal system. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology, 3rd ed*, pages 159–181. Cambridge University Press, New York, NY, US.

Dedovic, K., Renwick, R., Mahani, N. K., Engert, V., Lupien, S. J., and Pruessner, J. C. (2005). The montreal imaging stress task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience*, 30(5):319.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

der Zwaag, M. D. V., Janssen, J. H., and Westerink, J. H. D. M. (2013). Directing physiology and mood through music: Validation of an affective music player. *IEEE Transactions on Affective Computing*, 4(1):57–68.

Diehl, J. J., Schmitt, L. M., Villano, M., and Crowell, C. R. (2012). The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. *Research in Autism Spectrum Disorders*, 6(1):249–262.

Drachen, A., Nacke, L. E., Yannakakis, G., and Pedersen, A. L. (2010). Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, Sandbox '10, pages 49–54, New York, NY, USA. ACM.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., and Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied Ergonomics*, 52:95 – 103.

Fleuret, F. (2004). Fast binary feature selection with conditional mutual information. *J. Mach. Learn. Res.*, 5:1531–1555.

for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, S. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, 49(8):1017–1034.

Freeman, C., Kulić, D., and Basir, O. (2015). An evaluation of classifier-specific filter measure performance for feature selection. *Pattern Recognition*, 48(5):1812 – 1826.

Ghaderyan, P. and Abbasi, A. (2016). An efficient automatic workload estimation method based on electrodermal activity using pattern classifier combinations. *International Journal of Psychophysiology*, 110:91 – 101.

Giakoumis, D., Tzovaras, D., and Hassapis, G. (2013). Subject-dependent biosignal features for increased accuracy in psychological stress detection. *International Journal of Human-Computer Studies*, 71(4):425 – 439.

Giakoumis, D., Tzovaras, D., Moustakas, K., and Hassapis, G. (2011). Automatic recognition of boredom in video games using novel biosignal moment-based features. *IEEE Transactions on Affective Computing*, 2(3):119–133.

Giullian, N., Ricks, D., Atherton, A., Colton, M., Goodrich, M., and Brinton, B. (2010). Detailed requirements for robots in autism therapy. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 2595–2602.

Gold, L. H. (2014). DSM-5 and the assessment of functioning: the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *The Journal of the American Academy of Psychiatry and the Law*, 42(2):173–181.

Gomez-Herrero, G., Clercq, W. D., Anwar, H., Kara, O., Egiazarian, K., Huffel, S. V., and Paesschen, W. V. (2006). Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel. In *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, pages 130–133.

Greco, A., Valenza, G., Citi, L., and Scilingo, E. P. (2017). Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal*, 17(3):716–725.

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., and Citi, L. (2016). cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 63(4):797–804.

Gupta, A., Joshi, S. D., and Prasad, S. (2005). A new approach for estimation of statistically matched wavelet. *IEEE Transactions on Signal Processing*, 53(5):1778–1793.

Halligan, P. and Wade, D. (2005). *The Effectiveness of Rehabilitation for Cognitive Deficits*. Oxford University Press.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage Publications Sage CA: Los Angeles, CA.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183.

He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 1322–1328.

Healey, J. A. and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166.

Henriques, R. and Paiva, A. (2014). Learning effective models of emotions from physiological signals: The seven principles. In da Silva, H. P., Holzinger, A., Fairclough, S., and Majoe, D., editors, *Physiological Computing Systems*, pages 137–155, Berlin, Heidelberg. Springer Berlin Heidelberg.

Hernández, J., Morris, R. R., and Picard, R. W. (2011). *Call Center Stress Recognition with Person-Specific Models*, pages 125–134. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hjorth, B. (1970). Eeg analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306 – 310.

Horlings, R., Datcu, D., and Rothkrantz, L. J. M. (2008). Emotion recognition using brain activity. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, CompSysTech '08, pages 6:II.1–6:1, New York, NY, USA. ACM.

Hosseini, S. A. and Khalilzadeh, M. A. (2010). Emotional stress recognition system using eeg and psychophysiological signals: Using new labelling process of eeg signals in emotional stress state. In *2010 International Conference on Biomedical Engineering and Computer Science*, pages 1–6.

Hosseini, S. A., Khalilzadeh, M. A., Naghibi-Sistani, M. B., and Niazmand, V. (2010). Higher order spectra analysis of eeg signals in emotional stress states. In *2010 Second International Conference on Information Technology and Computer Science*, pages 60–63.

Höller, Y., Bergmann, J., Thomschewski, A., Kronbichler, M., Höller, P., Crone, J. S., Schmid, E. V., Butz, K., Nardone, R., and Trinka, E. (2013). Comparison of eeg-features and classification methods for motor imagery in patients with disorders of consciousness. *PLOS ONE*, 8(11):1–15.

Jenke, R. (2015). *Static and Dynamic Methods for Emotion Recognition from Physiological Signals*. PhD thesis, Technische Universität München, München.

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 5(3):327–339.

Jennings, J. R. and Gianaros, P. J. (2007). Methodology. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G., editors, *Handbook of Psychophysiology*, pages 812–833. Cambridge University Press, third edition. Cambridge Books Online.

Junli, L., Yanqin, H., Ping, W., and Gang, C. (2007). A novel method for the determination of the wavelet denoising threshold. In *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, pages 713–716.

Kamath, C. (2013). Teager energy based filter-bank cepstra in eeg classification for seizure detection using radial basis function neural network. *ISRN Biomedical Engineering*, 2013:9.

Kanamori, M., Suzuki, M., Oshiro, H., Tanaka, M., Inoguchi, T., Takasugi, H., Saito, Y., and Yokoyama, T. (2003). Pilot study on improvement of quality of life among elderly using a pet-type robot. In *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*, volume 1, pages 107–112 vol.1.

Kim, J. and André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083.

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

Kozima, H., Nakagawa, C., and Yasuda, Y. (2007). Children–robot interaction: a pilot study in autism therapy. In von Hofsten, C. and Rosander, K., editors, *From Action to Cognition*, volume 164 of *Progress in Brain Research*, pages 385–400. Elsevier.

Krishnaveni, V., Jayaraman, S., Anitha, L., and Ramadoss, K. (2006). Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients. *Journal of Neural Engineering*, 3(4):338–346.

Kurniawan, H., Maslov, A. V., and Pechenizkiy, M. (2013). Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 209–214.

Laparra-Hernández, J., Belda-Lois, J., Medina, E., Campos, N., and Poveda, R. (2009). EMG and GSR signals for evaluating user's perception of different types of ceramic flooring. *International Journal of Industrial Ergonomics*, 39(2):326 – 332.

Lazarus, R. S. (1991). *Emotion and Adaptation.* Oxford University Press, Oxford, New York.

Leite, I., Henriques, R., Martinho, C., and Paiva, A. (2013). Sensors in the wild: Exploring electrodermal activity in child-robot interaction. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 41–48.

Li, M. and Lu, B. L. (2009). Emotion classification based on gamma-band eeg. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1223–1226.

Li, M. and Narayanan, S. (2010). Robust ecg biometrics by fusing temporal and cepstral information. In *2010 20th International Conference on Pattern Recognition*, pages 1326–1329.

Lima, C. S., Tavares, A., H., J., J., M., and Barbos, D. (2010). Non-Stationary Biosignal Modelling. In Campolo, D., editor, *New Developments in Biomedical Engineering*. InTech.

Liu, C., Conn, K., Sarkar, N., and Stone, W. (2008). Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Transactions on Robotics*, 24(4):883–896.

Marti, P., Fano, F., Palma, V., Pollini, A., Rullo, A., and Shibata, T. (2005). *Symposium on Robot Companion Hard Problem and Open Challenges in Human-Robot Interaction*, volume Proc. AISB'05, page 64–73. Society of the Study of Artificial Intelligence and the Simulation of Behaviour (AISB).

Mauss, I. B. and Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & emotion*, 23(2):209–237.

McKeown, G., Valstar, M., Cowie, R., Pantic, M., and Schroder, M. (2012). The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 3(1):5–17.

McMahan, T., Parberry, I., and Parsons, T. D. (2015). Evaluating player task engagement and arousal using electroencephalography. *Procedia Manufacturing*, 3:2303 – 2310. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.

Meyer, P. E. and Bontempi, G. (2006). On the use of variable complementarity for feature selection in cancer classification. In Rothlauf, F., Branke, J., Cagnoni, S., Costa, E., Cotta, C., Drechsler, R., Lutton, E., Machado, P., Moore, J. H., Romero, J., Smith, G. D., Squillero, G., and Takagi, H., editors, *Applications of Evolutionary Computing*, pages 91–102, Berlin, Heidelberg. Springer Berlin Heidelberg.

Ming-Zher Poh, Loddenkemper, T., Swenson, N. C., Goyal, S., Madsen, J. R., and Picard, R. W. (2010). Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor. pages 4415–4418. IEEE.

Miranda-Correa, J. A., Abadi, M. K., Sebe, N., and Patras, I. (2017). Amigos: A dataset for affect, personality and mood research on individuals and groups.

Molavi, B. and Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement*, 33(2):259–270.

Montgomery, J. M., Newton, B., and Smith, C. (2008). Test Review: Gilliam, J. (2006). GARS-2: Gilliam Autism Rating Scale–Second Edition. Austin, TX: PRO-ED. *Journal of Psychoeducational Assessment*, 26(4):395–401.

Morris, J. D. (1995). Observations: SAM: The Self-Assessment Manikin; An Efficient Cross-Cultural Measurement of Emotional Response. *Journal of Advertising Research*, 35(8):63–38.

Moyle, W., Cooke, M., Beattie, E., Jones, C., Klein, B., Cook, G., and Gray, C. (2013). Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial. *Journal of Gerontological Nursing*, 39(5):46–53.

Moyle, W., Jones, C., Cooke, M., O'Dwyer, S., Sung, B., and Drummond, S. (2014). Connecting the person with dementia and family: a feasibility study of a telepresence robot. *BMC Geriatrics*, 14(1):7.

Nacke, L., Drachen, A., and Göbel, S. (2010). Methods for evaluating gameplay experience in a serious gaming context. *International Journal of Computer Science in Sport*, 9(2):1–12.

Nason, G. P. and Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. pages 281–300. Springer-Verlag.

Nihira, K., Leland, H., Lambert, N. M., Pro-Ed (Firm), American Association on Mental Retardation, and American Association on Mental Deficiency (1993). *ABS-RC:2 AAMR Adaptive Behavior Scale: residential and community*. Pro-Ed, Austin, Tex.

Organization, W. H. (2001). *The World health report : 2001 : Mental health : new understanding, new hope*. World Health Organization (WHO), Geneva, Switzerland.

O'Connell, R. G., Bellgrove, M. A., Dockree, P. M., Lau, A., Fitzgerald, M., and Robertson, I. H. (2008). Self-alert training: Volitional modulation of autonomic arousal improves sustained attention. *Neuropsychologia*, 46(5):1379 – 1390.

Peñaloza-Salazar, C., Gutiérrez-Maldonado, J., Ferrer-García, M., et al. (2015). Cognitive mechanisms underlying Armoni: A computer-assisted cognitive training programme for individuals with intellectual disabilities. *Anales de Psicología / Annals of Psychology*, 32(1):115–124.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., and Pioggia, G. (2015). Autism and social robotics: A systematic review. *Autism Research*, 9(2):165–183.

Petrantonakis, P. C. and Hadjileontiadis, L. J. (2010). Emotion recognition from eeg using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):186–197.

Piacentini, R. (2004). *Emotions at fingertips, "revealing individual features in galvanic skin response signals*. PhD thesis, Universit degli studi di Roma, Roma.

Pino, M., Boulay, M., Jouen, F., and Rigaud, A. S. (2015). "Are We Ready for Robots That Care for Us?" Attitudes and Opinions of Older Adults Towards Socially Assistive Robots. *Frontiers in Aging Neuroscience*, 7(141).

Rabbitt, S. M., Kazdin, A. E., and Scassellati, B. (2015). Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical Psychology Review*, 35:35–46.

Rani, P., Liu, C., Sarkar, N., and Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, 9(1):58–69.

Ricks, D. and Colton, M. (2010). Trends and considerations in robot-assisted autism therapy. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 4354–4359.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Sahli, A., Trecourt, P., and Robin, S. (1997). One sided acceptance sampling by variables: the case of the laplace distribution. *Communications in Statistics - Theory and Methods*, 26(11):2817–2834.

Sano, A. and Picard, R. W. (2013). Stress recognition using wearable sensors and mobile phones. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 671–676.

Schmidt, L. A. and Trainor, L. J. (2001). Frontal brain electrical activity (eeg) distinguishes valence and intensity of musical emotions. *Cognition and Emotion*, 15(4):487–500.

Scholtz, J. (2002). Evaluation methods for human-system performance of intelligent systems. Technical report, DTIC Document.

Shimomura, Y., Yoda, T., Sugiura, K., Horiguchi, A., Iwanaga, K., and Katsuura, T. (2008). Use of frequency domain analysis of skin conductance for evaluation of mental workload. *Journal of PHYSIOLOGICAL ANTHROPOLOGY*, 27(4):173–177.

Shukla, J. (2015). Using humanoid robots to convey rehabilitation therapies to disabled people. In *2nd URV Doctoral Workshop in Computer Science and Mathematics*, pages 63–65, Tarragona, Spain. Universitat Rovira i Virgili. ISBN: 978-84-8424-399-1.

Shukla, J. (2017). Employing socially assistive robotics to empower cognitive stimulation. In *5th JIPI: Jornada d'Investigadors Predoctorals Interdisciplinària*, page 6, Barcelona, Spain. Universitat de Barcelona.

Shukla, J., Barreda-Ángeles, M., Oliver, J., Nandi, G. C., and Puig, D. (2018a). Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Transactions on Affective Computing*, Submitted.

Shukla, J., Barreda-Ángeles, M., Oliver, J., and Puig, D. (2016a). Muderi: Multimodal database for emotion recognition among intellectually disabled individuals. In *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016*, pages 264–273. Springer International Publishing, Cham.

Shukla, J., Barreda-Ángeles, M., Oliver, J., and Puig, D. (2018b). Efficient wavelet-based artifact removal for electrodermal activity in real-world applications. *Biomedical Signal Processing and Control*, 42C:45 – 52.

Shukla, J., Barreda-Ángeles, M., Oliver, J., and Puig, D. (2017a). Effectiveness of socially assistive robotics during cognitive stimulation interventions: Impact on caregivers. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 62–67.

Shukla, J., Cristiano, J., Amela, D., Anguera, L., Vergés-Llahí, J., and Puig, D. (2015). Social robotics: 7th international conference, icsr 2015, paris, france, october 26-30, 2015. chapter A Case Study of Robot Interaction Among Individuals with Profound and Multiple Learning Disabilities, pages 613–622. Springer International Publishing, Cham.

Shukla, J., Cristiano, J., Anguera, L., Vergés-Llahí, J., and Puig, D. (2016b). Robot 2015: Second iberian robotics conference: Advances in robotics, volume 2. chapter A Comparison of Robot Interaction with Tactile Gaming Console Stimulation in Clinical Applications, pages 435–445. Springer International Publishing, Cham.

Shukla, J., Cristiano, J., Oliver, J., and Puig, D. (2018c). Robot assisted interventions for individuals with intellectual disabilities: Impact on users and caregivers. *International Journal of Social Robotics*, Submitted.

Shukla, J., Oliver, J., and Puig, D. (2017b). A computer-implemented method for the measurement of human emotion of a subject and a method for filtering an eda signal. European Patent Application Number EP17382661, Submitted.

Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1):140 – 164.

Sneddon, I., McRorie, M., McKeown, G., and Hanratty, J. (2012). The belfast induced natural emotion database. *IEEE Transactions on Affective Computing*, 3(1):32–41.

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55.

Spector, A., Gardner, C., and Orrell, M. (2011). The impact of cognitive stimulation therapy groups on people with dementia: Views from participants, their carers and group facilitators. *Aging & Mental Health*, 15(8):945–949.

Spector, A., Orrell, M., and Woods, B. (2010). Cognitive stimulation therapy (cst): effects on different areas of cognitive function for people with dementia. *International Journal of Geriatric Psychiatry*, 25(12):1253–1258.

Standen, P., Brown, D., Roscoe, J., et al. (2014). Engaging Students with Profound and Multiple Disabilities Using Humanoid Robots. In Stephanidis, C. and Antona, M., editors, *Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge*, volume 8514 of *Lecture Notes in Computer Science*, pages 419–430. Springer International Publishing.

Swangnetr, M. and Kaber, D. B. (2013). Emotional state classification in patient-robot interaction using wavelet analysis and statistics-based feature selection. *IEEE Transactions on Human-Machine Systems*, 43(1):63–75.

Tapus, A., Mataric, M., and Scasselati, B. (2007). Socially assistive robotics [grand challenges of robotics]. *Robotics Automation Magazine, IEEE*, 14(1):35–42.

Tapus, A., Tapus, C., and Mataric, M. (2009a). The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In *Rehabilitation Robotics, 2009. ICORR 2009. IEEE International Conference on*, pages 924–929.

Tapus, A., Tapus, C., and Mataric, M. J. (2009b). The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In *2009 IEEE International Conference on Rehabilitation Robotics*, pages 924–929.

Taylor, S. A., Jaques, N. M., Chen, W., Fedor, S., Sano, A., and Picard, R. W. (2015). Automatic identification of artifacts in electrodermal activity data. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (5):1934–7.

Ünal, Ö., Özcan, Ö., Öner, Ö., Akcakin, M., Aysev, A., and Deda, G. (2009). Eeg and mri findings and their relation with intellectual disability in pervasive developmental disorders. *World Journal of Pediatrics*, 5(3):196–200.

Vos, T., Barber, R. M., Bell, B., and Murray, . C. J. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, 386(9995):743–800.

Wagemaker, E., Dekkers, T. J., Rentergem, J. A. A. V., Volkers, K. M., and Huizenga, H. M. (2017). Advances in mental health care: Five n = 1 studies on the effects of the robot seal paro in adults with severe intellectual disabilities. *Journal of Mental Health Research in Intellectual Disabilities*, 10(4):309–320.

Wang, J. and Gong, Y. (2008). Recognition of multiple drivers emotional state. In *2008 19th International Conference on Pattern Recognition*, pages 1–4.

Wilson, G. F. and Fisher, F. (1995). Cognitive task classification based upon topographic eeg data. *Biological Psychology*, 40(1):239 – 250. EEG in Basic and Applied Settings.

Worthington, A. (2005). Rehabilitation of executive deficits. *The Effectiveness of Rehabilitation for Cognitive Deficits*, page 257–268.

Yakowitz, S. (1994). Time series analysis of higher order crossings (b. kedem). *SIAM Review*, 36(4):680–682.

Yang, H. H. and Moody, J. (1999). Data visualization and feature selection: New algorithms for nongaussian data. In *in Advances in Neural Information Processing Systems*, pages 687–693. MIT Press.

Zhai, J., Barreto, A. B., Chin, C., and Li, C. (2005). Realization of stress detection using psychophysiological signals for improvement of human-computer interactions. In *Proceedings. IEEE SoutheastCon, 2005.*, pages 415–420.

Appendix A

Questionnaire for the evaluation of robotic interaction effects

Description

As explained in section 4.1.1, there is no availability of any method or scale for the evaluation of robotic interaction effects. Thus, we used a questionnaire in collaboration with an expert psychiatrist at FAM for such evaluations. This questionnaire was adapted from GARS-2 Montgomery et al. (2008), WHODAS 2.0 Gold (2014) and ABS-RC: 2 Nihira et al. (1993) and consists of 12 questions. The Gilliam Autism Rating Scale-Second Edition (GARS-2) is a supplementary screening tool for individuals suffering with autism spectrum disorders and is scored on a scale of $0-3$. ABS-RC: 2 is a method to assess adaptive behavior of mentally handicapped persons and is scored on a scale of $0-4$. World Health Organization Disability Assessment Schedule 2 (WHODAS 2.0) is a tool for assessment of global functioning and impairment and is scored on a scale of $0-5$. Since all the ratings were evaluated on different scores, rating scores were converted to percentages. The questions presented below were selected to evaluate the social interaction and communication intents of the patient while he/she is interacting with other people or with the robot during the activity.

1. Does individual stare or look unhappy or unexcited when praised, humored, or entertained?

2. Does individual generally understand what people/robot say?

3. Does individual avoid eye contact (looks away when someone looks at him/her).

4. Does individual stare or look unhappy or unexcited when praised, humored, or entertained.

5. Does individual is non-imitative of other people when playing.
6. Does individual behave in an unreasonably fearful, frightenend manner.
7. Does individual look through people (ie, shows no recognition that a person is present).
8. Does individual laugh, giggle, cry inappropriately.
9. Does individual do certain things repetitively, ritualistically.
10. Does individual respond negatively or with temper tantrums when given commands, requests, or directions.
11. Does individual line up objects in precise, orderly fashion and becomes upset when the order is disturbed.
12. Is individual able to concentrate on doing something for ten minutes?