



J U S T A S Š A L K E V I Č I U S

**V I R T U A L R E A L I T Y F O R
A N X I E T Y R E C O G N I T I O N
U S I N G P H Y S I O L O G I C A L
S I G N A L S – F R A M E W O R K ,
M E T H O D S A N D
E V A L U A T I O N**

D O C T O R A L D I S S E R T A T I O N

K a u n a s
2 0 2 0

KAUNAS UNIVERSITY OF TECHNOLOGY

JUSTAS ŠALKEVIČIUS

VIRTUAL REALITY FOR ANXIETY
RECOGNITION USING PHYSIOLOGICAL
SIGNALS – FRAMEWORK, METHODS AND
EVALUATION

Doctoral dissertation
Technological Sciences, Informatics Engineering (T 007)

2020, Kaunas

This doctoral dissertation was prepared at Kaunas University of Technology, Faculty of Informatics during the period of 2015–2019.

Scientific Supervisor:

Prof. dr. Robertas DAMAŠEVIČIUS (Kaunas University of Technology, Technological Sciences, Informatics Engineering, T 007).

Doctoral dissertation has been published in:

<http://ktu.edu>

Editor:

Armandas Rumšas (Publishing Office “Technologija”)

KAUNO TECHNOLOGIJOS UNIVERSITETAS

JUSTAS ŠALKEVIČIUS

VIRTUALIOJI REALYBĖ NERIMO
ATPAŽINIMUI NAUDOJANT FIZIOLOGINIUS
SIGNALUS: SISTEMA, METODAI IR
VERTINIMAS

Daktaro disertacija
Technologijos mokslai, Informatikos inžinerija (T 007)

2020, Kaunas

Disertacija rengta 2015–2019 metais Kauno technologijos universiteto Informatikos fakultete.

Mokslinis vadovas:

Prof. dr. Robertas DAMASEVICIUS (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija, T 007).

Interneto svetainės, kurioje skelbiama disertacija, adresas:

<http://ktu.edu>

Redagavo:

Armandas Rumšas (leidykla “Technologija”)

CONTENTS

LIST OF FIGURES	8
LIST OF TABLES	10
ABBREVIATIONS	11
INTRODUCTION	13
1.1 The relevance of the work	13
1.2 Object of the work	14
1.3 Aim of the work	14
1.4 Tasks of the work	15
1.5 Scientific novelty	15
1.6 Practical value	16
1.7 Thesis statements	16
1.8 Scientific approval	16
1.9 Thesis structure	17
2 BACKGROUND OF VIRTUAL REALITY EXPOSURE THERAPY	19
2.1 Anxiety disorders and their treatment	19
2.2 Virtual reality exposure therapy	20
2.3 Architectures and models of VR systems	24
2.4 Affective computing in virtual environments	26
2.5 Recognition of stress and anxiety	28
2.6 Psychophysiological signals for anxiety detection systems	29
2.7 Psychophysiological feature extraction	30
2.8 Anxiety scales	33
2.9 Psychophysiological signal processing, feature extraction and classification methods for recognition of the emotional (affective) state	34
2.10 Conclusions of the Chapter	37
3 MODELS AND METHODS FOR VRET WITH ANXIETY RECOGNITION	38
3.1 Analysis of the requirements for VRET system	38
3.2 Virtual Reality as a Service (VRaaS) model	40
3.3 Proposed VRET system	44
3.3.1 VR headset selection	45

3.3.2	Designing VR scenes for public speaking anxiety	48
3.3.3	Anxiety recognition in VRET system.....	51
3.3.4	Software tools	51
3.4	Evaluation.....	52
3.5	Conclusions of the Chapter.....	53
4	VRET ANXIETY LEVEL CLASSIFICATION METHOD	55
4.1	Anxiety level classification for VRET	55
4.2	Signal normalization.....	56
4.3	Signal denoising	57
4.4	Windowing	57
4.5	Feature extraction	57
4.6	Anxiety level class assignment.....	59
4.7	VRET anxiety classification.....	60
4.8	Conclusions of the Chapter.....	61
5	EXPERIMENTS AND RESULTS	62
5.1	Participant demographics	62
5.2	Flow of the experiment.....	62
5.3	VRET stimuli.....	63
5.4	Psychophysiological signal collection through a wearable sensor	64
5.5	Anxiety level evaluation when using SUDS	65
5.6	Results	65
5.6.1	Collected psychophysiological signals.....	65
5.6.2	Class distribution	69
5.6.3	Validation methods.....	69
5.6.4	Signal window size evaluation	70
5.6.5	Feature evaluation.....	71
5.6.6	Psychophysiological signal evaluation and signal fusion.....	73
5.6.7	Leave one out subject validation	76
5.6.8	Comparison of results with other works.....	77
5.7	Future work	79
5.8	Conclusions of the Chapter.....	80
6	CONCLUSIONS	82
	REFERENCES	83

LIST OF PUBLICATIONS..... 95

LIST OF FIGURES

Figure 1. VR system design by Bekele <i>et al.</i> [57].....	24
Figure 2. VRET system architecture with difficulty levels by Dhiman <i>et al.</i> [58]	24
Figure 3. Architecture of online remote VRET system by Brinkman <i>et al.</i> [59]	25
Figure 4. Virtual exposure therapy architecture with personalized customization by Ben-Moussa <i>et al.</i> [60].....	25
Figure 5. Process of decomposing BVP into heart rate.....	30
Figure 6. VRaaS model for Virtual Reality therapy system.....	41
Figure 7. SaaS deployment for a virtual reality exposure therapy system	44
Figure 8. Minimal VRET system setup.....	46
Figure 9. Patient overview page with a session history.....	47
Figure 10. VRET control options	49
Figure 11. VRET environments created for the treatment of public speaking anxiety	50
Figure 12. Normalized GSR and skin temperature session values by experiment participant	57
Figure 13. Raw and filtered biofeedback signals – a) GSR and b) skin temperature	57
Figure 14. Anxiety level class assignment by SUDS value	60
Figure 15. Flow of the experiment. VRET is conducted in Step 6.....	63
Figure 16. VRET public speaking environment: a) backstage; b) stage in the virtual conference hall	64
Figure 17. Values reported on the subjective units of the distress scale	65
Figure 18. Subjective units of distress evaluation: a) those who agreed or strongly with the statement “While preparing for giving a speech, I feel tense and nervous”; b) all the remaining subjects	66
Figure 19. Differences in the SUDS values reported after a VRET session: a) based on the previous gaming experience; b) based on the previous VR experience	66
Figure 20. Normalized GSR and skin temperature session values by participant of experiment.....	68
Figure 21. Anxiety class distribution.....	69
Figure 22. Anxiety level prediction by using blood volume pulse (BVP), galvanic skin response (GSR), and skin temperature: accuracy dependency upon signal window length.....	70
Figure 23. Feature importance BVP (ten most relevant features are shown). 71	
Figure 24. Feature importance GSR (ten most relevant features are shown). 72	
Figure 25. Feature importance skin temperature (ten most relevant features are shown)	72
Figure 26. Feature importance signal fusion (ten most relevant features are shown)	73
Figure 27. GSR signal classification normalized confusion matrix	74

Figure 28. BVP signal classification normalized confusion matrix	74
Figure 29. Skin temperature signal classification normalized confusion matrix	75
Figure 30. Signal fusion (early) classification normalized confusion matrix.	76
Figure 31. Results of one-subject-out validation.....	77

LIST OF TABLES

Table 1. Comparison of VRETs across different studies.....	26
Table 2. Feature set of physiological signals.....	31
Table 3. Virtual reality systems with psychophysiological signal tracking....	36
Table 4. VRET system criteria and requirements	39
Table 5. Signals (specification) collected in proposed VRET system.....	51
Table 6. Features extracted from physiological signals.....	58
Table 7. Demographic information of the participants.....	62
Table 8. Subjective Units of distress scale	65
Table 9. Subjective signs of discomfort	67
Table 10. Elevated GSR and skin temperature feature values.....	68
Table 11. Anxiety level prediction accuracy by signal type and window size (best results are shown in bold)	71
Table 12. Anxiety level prediction accuracy by signal type (when using 10x10-fold cross validation).....	76
Table 13. Comparison with laboratory induced stress recognition systems...	78
Table 14. Comparison with driving induced stress recognition systems.....	79
Table 15. Comparison with public speaking induced stress recognition systems	79

ABBREVIATIONS

3D – third dimensional

AC – affective computing

API – application programming interface

BCI – brain-computer interface

BVP – blood volume pulse

CBT – cognitive behavioral therapy

ECG – electrocardiogram

EEG – electroencephalogram

ELM – extreme learning machine

ET – exposure therapy

FPR – false positive rate

GRNN – general regression neural network

GSR – galvanic skin response

HCI – human-computer interaction

IPQ – Igroup presence questionnaire

LDA – linear discriminant analysis

LVQ – learning vector quantization

MA – moving average

OET – one session exposure therapy

PCA – principal component analysis

PRCA-24 – personal report of communication apprehension

PSAS – public speaking anxiety scale

PTSD – post-traumatic stress disorder

REST – representational state transfer

RR interval – the time between QRS complexes

QRS – the combination of three of the graphical deflections seen on ECG

SATI – speech anxiety thoughts inventory

SUDS – subjective units of distress scale

SVM – support vector machine

SaaS – software as a service

VR – virtual reality

VRET – virtual reality exposure therapy

WHO – World Health Organization

INTRODUCTION

1.1 The relevance of the work

Anxiety disorders, such as social phobia, public speaking anxiety, fear of flights, post-traumatic stress disorder, and others can severely limit and impact patients' daily life activities. Anxiety usually leads to such symptoms as tension, panic, and irritability. The terms *stress* and *anxiety* are often used interchangeably as they play an important role in human psychology. Stress (anxiety) is a normal bodily reaction to stressors (stimuli) from the environment. However, prolonged or abnormal anxiety can lead to dysfunction and distress. Thus, it is important to investigate the already existing and new methods and tools that could be used for anxiety treatment.

Virtual reality exposure therapy (VRET) can help solve some of these problems by placing the patients in a simulated 3D environment where they can gradually be exposed to the feared stimuli and learn to control their reactions. These environments can be preplanned and arranged before the treatment sessions. Additionally, virtually simulated environments add ability for the patient to repeat them in the same scenario for multiple times. Moreover, VRET sessions can be organized confidently and privately within the walls of a psychology clinic. They also have a benefit of cost saving in the long run, as the initial virtual reality system development can be expensive, but, once finished, the same 3D environments and treatment scenarios can be used for multiple patients across multiple psychology clinics. However, they are several barriers and limitations for a VRET-based treatment. Firstly, the adaptation of VRET systems is slowed by the need for VR training for the therapists, hardware requirements (a VR headset, a powerful PC to run VR software), lack of education about VR advantages, and others [1]. Secondly, it is important to note that the surveys of VRET studies showed that VRET is neither less, nor more efficient than the traditional therapy, nor does it improve the dropout rates [2]. Although, in the cases where it is difficult to arrange the appropriate treatment environments, VRET can serve as a great alternative.

As previous researchers have shown, the virtual reality exposure therapy (VRET) can be used as a tool during therapist-led anxiety disorder interventions. One of the potential strengths of VRET systems is that they provide an opportunity for a psychologist to interact with virtual 3D environments and change therapy scenarios according to the individual patient's needs. However, the design and the architecture of many VRET systems is standalone by their nature. This means that the therapist can only observe the patient's actions in a preplanned scenario, and has no ability to change the virtual environment to better suit the specific patient.

Dynamic interaction with the VRET system for the therapist can open up additional treatment options. These interactions can provide ability to control virtual avatars and virtual crowds, change their behavior and reactions based on the patient's actions during the session, and create either less or more stressful experiences according to the patient's progression. Thus, the same scenario can be used multiple times with unexpected and new adaptations. Additionally, the

environment itself can be changed to reflect daily or nightly scenarios, or, in the case of driving anxiety, traffic can be either light or really heavy.

Moreover, stressful or feared situations can be initiated by the therapist on the required scale of intensity, like turbulence during a flight for flight anxiety, car incidents for driving anxiety, or a computer crash during a presentation for public speaking anxiety. Finally, various audio-based stimuli can be integrated and triggered by the therapist at appropriate times. These can be either sounds for an environment, such as noisy traffic and drivers yelling, or audio cues related to panic attacks, like extensive heartbeats or heavy breathing.

However, in order to achieve higher efficiency, the patient's anxiety level should be tracked throughout the VRET session. This would allow the therapist to evaluate the current state of the patient more easily and to change the course of the virtual environment and the scenario. Therefore, in order to fully use all the advantages provided by the VRET system, a mental stress detection system is needed.

Anxiety recognition systems are widely researched and used in various scenarios. Human body reactions to stress can be observed by measuring bodily reactions controlled by the autonomic nervous system (ANS). ANS can be separated into two major systems: the parasympathetic nervous system and the sympathetic nervous system. The parasympathetic system is responsible for calming reactions and returning the nerves to the neutral state, while the activation of the sympathetic system is activating the body reaction to the potential threat. These changes are reflected through psychophysiological signals that can be collected with wearable non-invasive biofeedback sensors. Such signals as the blood volume pressure (BVP), galvanic skin response (GSR), respiratory rate (RV), heart rate (HR), electroencephalogram (EEG), electromyography (EMG) and skin temperature can be processed to recognize anxiety. This requires processing of various signals, feature extraction, feature selection, and classification methods to be implemented.

The anxiety recognition component integration into an interactive VRET system could lead to novel advances and improvements in this kind of therapy. Therefore, this work focuses its efforts onto designing the architecture and methods for this kind of system.

1.2 Object of the work

The object of this research is an anxiety-aware Virtual-Reality-as-a-Service (VRaaS) model.

1.3 Aim of the work

The aim of the work is to develop a virtual reality model that could be used as a tool for inducing the desired human affective states by using virtual reality stimuli and recognizing affective states (anxiety) by using psychophysiological signals.

1.4 Tasks of the work

The main task of this study is the design and investigation of adaptive VRET system with an integrated anxiety recognition component, which includes the following subtasks:

1. Analyze state-of-the-art VRET systems, their models, architectures, and advanced usage capabilities, and stress detection frameworks which employ psychophysiological signal analysis.
2. Propose an adaptive design of a VRET system based on a Virtual-Reality-as-a-Service (VRaaS) model.
3. Develop practical implementation of the proposed anxiety recognition component with an ability for the therapist to adapt the exposition therapy session scenario in real-time.
4. Evaluate psychophysiological signal processing and feature extraction techniques for the anxiety recognition component working within the VRET system.

1.5 Scientific novelty

The work achieved these novel results:

1. The proposed VRaaS model enables an adaptable VRET system with an integrated anxiety recognition component. This system provides ability for the therapist to adapt virtual environments, change reactions of virtual avatars, and trigger various situations during treatment sessions in order to personalize the treatment session for a specific patient.
2. The suggested VRET system features an integrated four-class anxiety recognition model, while the work of previous researches mostly focused on recognizing anxiety in two or three classes. This model provides the ability for the therapist to recognize the patient's anxiety during VRET at more distinct levels (low, mild, moderate, and high anxiety). Based on this capability to recognize anxiety in a more precise manner, the therapist has more information on how to adapt the current virtual environment scenario for the patient in real-time (to lower or to increase the intensity of the therapy).

1.6 Practical value

Most of the VRET system is based on the standalone applications running on a PC. This thesis describes an architecture of a cloud-based VRET solution that can be used across multiple psychology clinics and with a lightweight setup.

Additionally, the VRET system designed and described in this thesis provides capabilities for the therapist to actively participate during the session. By using the implemented VRET system, the therapist can change and influence the course of the therapy scenario and adapt it to the patients' needs.

Moreover, we design a psychophysiological signal processing and feature extraction component for anxiety recognition during the session. This would potentially help the therapist to measure more accurately the condition of the patient during the VRET session, thus allowing better adaptation of the VR scenario.

Additionally, the proposed VRET system with the ability to measure and recognize the patient's anxiety was supported by the *European Regional Development Fund* within the "Intelektas. Bendri mokslo-verslo projektai" program. Therefore, the implemented system has already been used in psychology clinics to conduct clinical trials.

This thesis shows the fundamental knowledge of VRET system design, feature extraction, and classification of patients' anxiety during VRET sessions when using machine learning methods. Finally, it presents an experiment with 30 participants for the treatment of public speaking anxiety conducted with a VRET system designed by the author of the thesis.

1.7 Thesis statements

1. A VRET system based on the Virtual-Reality-as-a-Service (VRaaS) model enables the development of an adaptive, affordable, mobile, and simple VRET toolkit for a psychology clinic.
2. The adaptive and interactive design of VRET systems allows the therapist to change the VR environments during the treatment sessions in real-time, thus providing more capabilities of personalized affective disorder analysis and treatment.
3. A VRET anxiety recognition system that uses psychophysiological signals collected through wearable sensors can be integrated within the bounds of the VRET system and is capable to recognize the patient's anxiety at four levels (low, mild, moderate, and high anxiety).

1.8 Scientific approval

This dissertation is partly based on the data and research experience gathered and described in the following peer-reviewed journal articles:

1. **Šalkevičius, Justas**, Miškinytė, Audronė, & Navickas, Lukas. (2019). Cloud based virtual reality exposure therapy service for public speaking anxiety. *Information*, 10(2:62), 1-15.
2. **Šalkevičius, Justas**, Robertas Damaševičius, Rytis Maskeliūnas, & Ilona Laukienė. (2019). Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals. *Electronics*, 8(9), 1039.

This dissertation is partly based on the following peer-reviewed conference publications:

1. **Šalkevičius, Justas**, and Navickas, Lukas. (2018). Battling the fear of public speaking: Designing software as a service solution for a virtual reality therapy. *W-FiCloud 2018: 6th International Conference on Future Internet of Things and Cloud Workshops*, 6–8 August 2018, Barcelona, Spain, 209-213.
2. **Šalkevičius, Justas**, and Damaševičius, Robertas. Adapting Virtual Environments Based on Patients' Anxiety during Virtual Reality Exposure Therapy. *11th International Workshop on Data Analysis Methods for Software Systems*, Druskininkai, Lithuania, November 28-30, 2019 (2019): 79.

1.9 Thesis structure

This thesis consists of 6 chapters. Chapter 1 is an introduction to the thesis which summarizes the aim, novelty, objectives, and motivation for the work.

Chapter 2 begins by describing VRET based anxiety treatments and contains a thorough review of the potential adaptations, advantages, and limitations of the currently available VRET systems. We review their usages and state-of-the-art techniques that are being used to design such systems. Additionally, we review the mental stress and anxiety recognition frameworks used in various fields. Moreover, we are reviewing psychophysiological signal processing and feature extraction methods that can be used in designing and implementing the mental stress recognition framework. The purpose of this section is to introduce and familiarize the reader with the field of VRET and anxiety recognition systems that are using psychological signal analysis in order to correctly identify the subject's emotional states.

Chapter 3 presents our suggested VRaaS model which enables an interactive VRET system with an ability to adapt virtual therapy scenarios based on the patient's anxiety. We carefully describe the thought process of developing the VRET system. We are starting from VR hardware selection, the architecture of the major components for our VR system, and then we are discussing the real-time interactions with 3D environments that our application provides for the therapist,

and, lastly, we are describing the VR scenes created for public speaking anxiety and the integration of anxiety.

Chapter 4 focuses on the methods that were used to implement the anxiety recognition model within our VRET system. We discuss the selected methods and techniques for psychophysiological signal processing, normalization, feature sets, and anxiety state labeling. We detail the design of the signal processing and feature the extraction methods.

Chapter 5 describes the experiment flow and the analysis of the classification results. We are discussing the classification windows size selection, signal feature selection, as well as the classification accuracy by using standalone signals and signal fusion. Lastly, we are performing independent subjects leave one out validation and comparing our results to similar studies in the field of anxiety recognition.

Chapter 6 contains conclusions and reviews the results achieved with our implemented VRET system with an anxiety recognition component tackled by our research.

2 BACKGROUND OF VIRTUAL REALITY EXPOSURE THERAPY

2.1 Anxiety disorders and their treatment

Anxiety disorders are reaching rates as high as 28.8%, according to World Health Organization statistics [3]. In many cases, anxiety or stress is a normal reaction of the organism in order to cope with unexpected events. However, abnormal reactions to simple daily stress stimuli can cause enduring anxiety. These abnormal reactions are classified as anxiety disorders, which include various phobias, panic, social anxiety, and generalized anxiety disorders, etc. More than 30% of the population has been suffering from an anxiety disorder at least once during their lifetime [4]. Anxiety can be caused by multiple causes, such as past experiences, exhaustion, job burnout, loneliness, abuse, etc., and it can lead to reduced quality of life and labor productivity at work [5]. Anxiety can severely impact a patient's daily life, trigger social communication problems, and diminish life quality. Moreover, anxiety disorders are usually chronic conditions and require prolonged treatment periods [6].

Anxiety disorders usually are noticeable in two dimensions: psychological and physical. Psychological symptoms are recognized as irritability, tension, disturbance, and restlessness, while physical reactions are involving the adrenal glands, the sympathetic system, and the production of hormones. Anxiety disorders are a part of mental illnesses and are manifested by various symptoms, like worry, avoidance of social contact, panic, fear, distress, and sleeplessness. These symptoms can be decoded by applying a variety of emotional stimuli, such as static and dynamic audio/video stimuli [7].

Anxiety disorders can be treated using cognitive behavioral therapy (CBT) [8]. For example, public speaking anxiety is usually treated by employing cognitive-behavioral therapy (CBT) combined with exposure therapy (ET). These therapy sessions are designed to let the patient experience their anxiety in a managed environment enabled by *in vivo* (real-world confrontation of fear stimuli) and an imagined exposure (when the patient is asked to vividly imagine and describe the fear stimulus) [9]. CBT is focusing on changing the patient's cognitive patterns and behaviors when faced with the feared stimuli and on learning new coping techniques. CBT is a popular treatment adapted by many therapists as interventions can be flexible and organized in many different forms. This includes group therapy sessions, individual interventions, or even self-therapy. The important distinction in CBT is how the patient confronts the feared object. As mentioned before, it can be done either *in vitro*, when the feared stimuli are imagined and vividly described, or *in vivo*, when the stimuli are faced in the real world. The *in vivo* method is a key feature of the exposure therapy, which provides the most efficient way of treating anxiety disorders. Therefore, in the clinical community, an exposure therapy was proven to be 'gold standard' for treating anxiety disorders [10].

Yet, ET raises several challenges for a successful treatment. First of all, it could be difficult to arrange a session in the appropriate environment: an airport for flight anxiety, traffic environment for driving anxiety, etc. Secondly, it could be difficult or almost impossible to organize confidential treatment sessions in public

spaces. Thirdly, if an *in vitro* session is arranged in the public space, it is not possible to control the environment. Fourthly, it could be even dangerous to put an anxious patient in the feared scenarios, like driving a vehicle through the city for driving anxiety. Lastly, even in the case of *in vivo* therapy, patients can have difficulties or be unwilling to imagine the object that causes anxiety.

2.2 Virtual reality exposure therapy

Sometimes it can be very costly to set up and manage a controlled environment for feared stimuli. Thus, cost-effectiveness is an important topic for anxiety disorder treatment when using exposure therapy [11]. New advances in virtual reality (VR) technologies and high-quality consumer-grade devices [8] have opened up new prospects for virtual reality exposure therapy (VRET). VR systems for exposure therapy can offer reusable 3D environments and interactive scenarios for the feared stimuli that can be difficult to replicate in real-life, such as a virtual battle environment for post-traumatic stress disorder (PTSD) [12], or a virtual flying environment [13]. Moreover, a recent survey of VRET studies showed that it could be equally effective as *in vivo* exposure therapy, and it even has a slightly lower dropout rate compared to the traditional treatment [14].

We can improve and analyze VRET systems better by tracking the patient's physiological signals [15][16]. The ability to track and recognize the patient's anxiety levels during VRET interventions can help to adapt and individualize the therapy. By implementing these capabilities in VRET systems, we can move them closer to achieving the goals of affective computing (AC) to detect and recognize human affective states or their disorders, such as depression [17]. AC has become an important branch of the field of human-computer interaction (HCI) [18]. Systems with affective computing capabilities can sense the emotions of the user. With this information, the system can adapt to the user by performing a predefined sequence of actions [19], such as changing the reactions of 3D avatars in the VR environment based on the patient's anxiety [20]. Many psychological cues can indicate the change of the user's affective state, such as facial expressions, posture, pulse, temperature, electrodermal activity (EDA), etc. In order to detect an emotional state, it is required to gather these emotional cues from biofeedback devices and then extract strong patterns and features. As many studies have shown before, the most advanced way to implement these requirements is to use machine learning models that can learn from features acquired from physiological signals, such as Blood Volume Pressure (BVP), Photoplethysmography (PPG), Heart Rate Variability (HRV), Galvanic Skin Response (GSR), Electroencephalography (EEG), Electromyography (EMG) and others [21][22][23][24].

Providing manageable environments could be a vast and expensive task, whereas the use of the virtual reality exposure therapy (VRET) can help to achieve the same desired effect by putting the patient into a simulated 3D environment [25]. At least 24 VRET studies were conducted and reviewed from 2012 to 2015 [26]. The authors of these VRET trials demonstrated that the use of virtual reality systems in the therapeutic setting is even more effective than the standard CBT session. Yet,

as more and more virtual reality headsets are reaching the market, it is increasingly important to analyze the impact of new devices on the quality of VRET sessions. Even simple hardware limitations, such as poor screen density or a poor VR sensor signal, can reduce the effectiveness of the therapy [27]. Therefore, more effort should be applied to investigate and measure the impact of newly developed VRET applications which enable the psychologist to interact with and adapt the 3D environment for the patient in real-time. Public speaking anxiety has been successfully treated by using various VRET implementations [28][29][30]. Yet, these implementations do not compare with a mobile VRET solution which can be easily accessed by wider audiences and which gives more VR scenario control options to the psychologist.

Additionally, it is important to select a correct physiological signal sensor device which could be comfortably used with the VR headset. A recent survey of physiology-based mental stress detection systems which use machine learning methods reported over 25 different wearable sensors which were employed to collect such signals [31]. However, for the data collection during VRET sessions, it is important that the VR device itself would not create additional discomfort and stress for the patient. Thus, wristband base biofeedback devices are preferable for VRET users. However, EEG data collection in combination with the VR systems is limited to custom-made setups and is not widely used with VR headsets.

Since Max North [32] pioneered the term *virtual reality therapy*, hundreds of studies have been dedicated to this field [33][34][35]. The most popular applications feature dental phobias, flight anxiety, public speaking anxiety, and fear of heights to PTSD.

An extensive review of multiple VRET studies was conducted by Botella *et al.* [36]. The authors noted that simulated VR scenarios can increase the effect of CBT sessions and can provide a faster result for the patient. The paper also states that VRET is limited by poor uptake by psychologists. This poor uptake is usually caused by the difficulty of setting up the VR hardware, lack of simple instructions for the user, and the cost of VRET systems.

Another systematic review of VRET applications was written by Carl *et al.* The authors noted that this kind of therapy is more effective than standard CBT treatments [37]. Moreover, this effectiveness can be seen across a variety of phobias.

An online VRET application was developed and reviewed by Levy *et al.* They used the online VR scenario for the treatment of acrophobia [38]. However, the authors stated that the extensive requirements imposed on the VR setup and the need for powerful PCs is halting the adoption of VRET solutions.

Suyanto *et al.* suggested that a mobile VRET solution could be improved by the *Kinect* sensor. Moreover, they were using the State-Trait Anxiety Inventory (STAI) scale in order to evaluate the reduction of self-observed anxiety levels before and after a VRET session [39].

Similarly, Shunnaq *et al.* used the Hamilton scale to measure the efficiency of their VRET solution [40]. The Subjective Units of Distress (SUD) scale was used by Rauch *et al.* in an extensive VRET study for PTSD (Posttraumatic stress disorder) [41].

A VR driving simulator was used by Zinzow *et al.* They have investigated the possibility of using a VRET system as a tool for the treatment of PTSD [42]. While some participants complained about the virtual reality environments not being realistic enough, the study showed a significant reduction in risky and aggressive driving.

Marquardt *et al.* proposed that VRET applications could be improved by using multisensory cues in order to increase engagement and to trigger a stronger emotional response [43]. Their system includes an olfactory device which can trigger a variety of smells aimed towards the user and which uses large fans to create a wind effect for acrophobia environments.

The effects of the 3D scene characteristics (space openness, tidiness, and color) in a virtual environment were analyzed by Christofi *et al.* Their research showed no significant increase of anxiety in messy virtual environments [44]. Contradicting reports were received from the participants that messy rooms increased their stress.

Poeschl *et al.* investigated the effects of a virtual crowd during a VRET session [29]. The facial expressions and gestures of the virtual audience were highlighted as important factors for increased stimuli.

Beidel *et al.* [45] investigated VRET efficiency for PTSD treatment by conducting a randomized controlled trial. During the study, more than 90 war veterans underwent VRET sessions with an effectiveness being evaluated mid- and post-treatment, and at 3/6 month follow-ups. The participants' anxiety levels throughout the VRET session were assessed by using the subjective unit of distress scale (SUDS) and were collected at 10 min intervals. This study confirmed a hypothesis that VRET resulted in a significant decrease in PTSD symptoms.

Automated VRET sessions for acrophobia were conducted by Freeman *et al.* [46]. In a period of 4 weeks, these sessions managed to significantly reduce the fear of heights for the study group. These results demonstrated the possibility of reducing therapy costs by using the VRET solution.

Manju *et al.* used a VR system in order to reduce social anxiety within a study group of kids with autism disorder [47]. Virtual environments provided the ability to repeat social interactions multiple times, which helped these patients to concentrate their minds into a single action.

Some researchers are already integrating the psychophysiological signal in their studies. However, in most cases, these observations are limited to statistical analysis with some initial steps to a fully fledged anxiety recognition system for VRET.

Another VRET study for anxiety disorder treatment was conducted by Diemer *et al.* [48]. The study included 40 patients with acrophobia and 40 healthy participants targeting the fear of height. The VR scenario increased the scene height every 2 minutes, and the participants' subjective fear was also assessed by using SUDS. Additionally, the authors collected GSR and HRV signals during their VRET sessions. The study found that the VR height scenario activated physiological fear reactions for both experiment groups.

Raghav *et al.* [49] suggested using VRET for dealing with dental phobias in adults. A randomized clinical trial was conducted among 30 patients. Two groups were separated, and one of them had a VRET session, while the other group was given informational pamphlets. A visual analogue scale for anxiety (VAS-A) measuring from 0 to 100 was used to record subjective fear, and participants' HRV signals were collected. The concluded study showed a significant reduction in the anxiety scores in the VRET group.

Freeman *et al.* used [46] immersive VR for treating the fear of heights. During the experiment, electrocardiogram (ECG), GSR and respiration data were collected. The experiment involved 61 participants who were involved in relaxation – stress – relaxation type tasks, each participant repeated this for 15 times in the interval of 4 weeks. When using this data, a binary classifier was trained for two classes: relaxed or stressed. The authors achieved 88% accuracy with the fused signal feature set using SVM.

The benefits of converting the traditional one-session exposure therapy (OST) to the VRET system are discussed by Miloff *et al.* During OST, the patient is gradually exposed to the stimuli for up to 3 hours, which can use up significant financial resources – while a VR system can help to reuse the same 3D environment for multiple patients [50]. Various authors reviewed the possibility of combining VRET systems with another stress management practice. Diaphragmatic breathing during a VRET session is suggested as a way to reduce stress [51]. A visual feedback system that could represent the current level of anxiety in the 3D environment was also investigated as an impactful way to control anxiety [52]. A VRET solution with a haptic feedback device that emulates heartbeat was described by Seol *et al.* [53].

Lindner *et al.* compared therapist-led and self-led VRET sessions conducted by using mobile devices [51]. Both groups in their study experienced a significant reduction of the public speaking anxiety. The advantages of mobile VRET solutions were also reviewed by Wiederhold *et al.* [54]. In their paper, the authors emphasized that, while the mobile solutions cannot offer high definition displays, they still can help to significantly reduce pain and anxiety.

On the other hand, the two recent systematic surveys of VRET studies indicated the following limitations of this kind of treatment [55][2]:

- VRET is neither less nor more effective than the traditional *in vivo* exposure therapy.
- VRET could be the preferential choice only in cases when it can be more accessible or preferable than the traditional alternatives.
- The use of VRET does not lower the attrition and dropout rates from the therapy course.

Sight impairments can be a major factor in the usefulness of VRET. Bun *et al.* separated their experiment participants into four groups – with and without the fear of heights and with and without sight impairments [56]. The authors observed that higher anxiety levels were induced by the VR stimuli for the group without sight impairment.

2.3 Architectures and models of VR systems

Bekele *et al.* suggested the VR architecture of standalone application where all the components – VR Task Engine (Unity), Eye tracker application and physiology signal modules – communicate through a unified network interface layer, although all the meta-data, such as physiology, eye tracking and Performance logs generated by components are stored and kept separately (**Figure 1**).

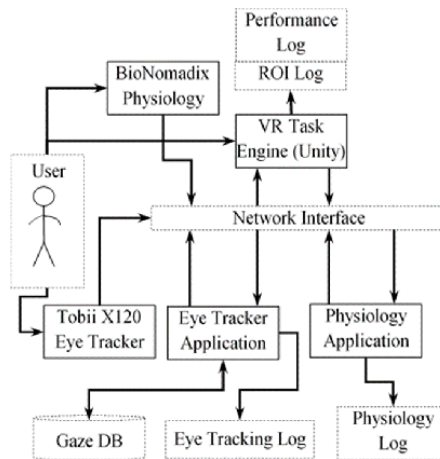


Figure 1. VR system design by Bekele *et al.* [57]

A more advanced VR system architecture was described by Dhiman *et al.* Their system offered three levels of difficulty for a VR-based task. A physiological data acquisition model would collect and analyze the user's physiological signals and would pass them to task the difficulty switching module. This module would react on the grounds of the predefined user's performance parameters based on previously analyzed signals and would switch the VR application either to a lower or to a higher difficulty setting controlled by the VR task module (**Figure 2**).

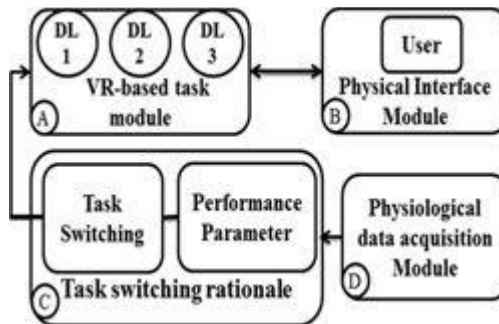


Figure 2. VRET system architecture with difficulty levels by Dhiman *et al.* [58]

The architecture of a remote VRET system with therapist led interactions was used by Brinkman *et al.* The therapist would spectate and listen to the

participant through a camera and a microphone. Based on the participant's behavior, the therapist would adapt and choose the dialogue from the dialogue database through the therapist's interface, and this dialog would be played for the participant (**Figure 3**).

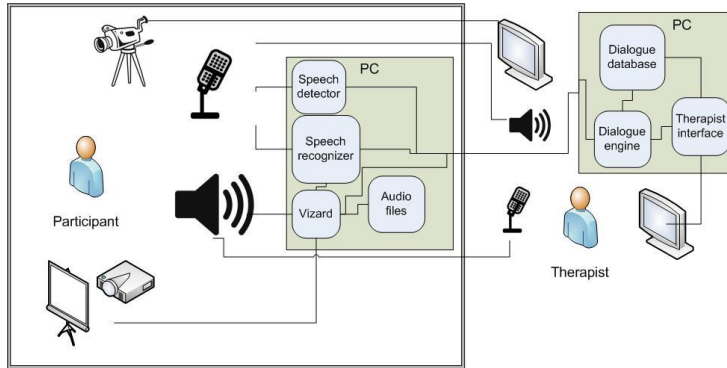


Figure 3. Architecture of online remote VRET system by Brinkman *et al.* [59]

Ben-Moussa *et al.* separated the virtual therapy system design into major components (**Figure 4**). Firstly, the user interface is delivered through an AR or VR headset with a smartphone application. This is followed by an intelligence and control module which is responsible for progress tracking, logging and patient behavior interpretation. These two components additionally rely upon pre-defined data including scenarios, workflows, and personalized customization, as well as perception technologies, such as psychological signals or audio analysis.

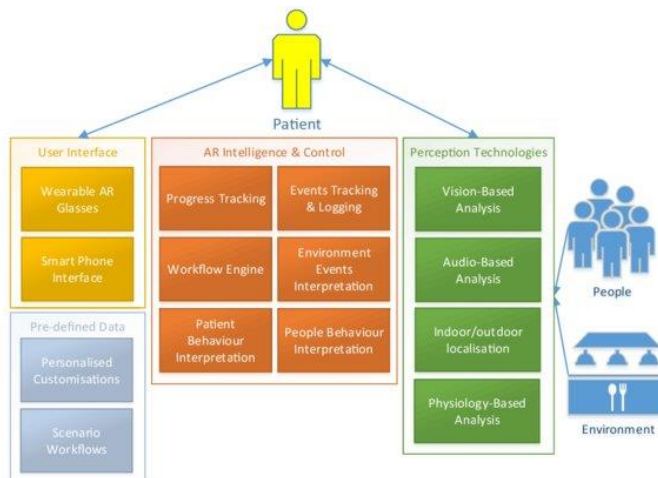


Figure 4. Virtual exposure therapy architecture with personalized customization by Ben-Moussa *et al.* [60]

We have compared these architectures based on the ability to adapt a VR scenario during the session, how the therapist can manage the VRET sessions, whether the affective state recognition is available, and how scalable the system can be to set it up and control it throughout multiple sites (**Table 1**).

Table 1. Comparison of VRETs across different studies.

	Bekele <i>et al.</i> [57]	Dhiman <i>et al.</i> [58]	Brinkman <i>et al.</i> [59]	Ben-Moussa <i>et al.</i> [60]
VR scenario adaptability	Low: there is no way to personalize the session	Medium: predefined difficulty levels	Medium: dialogue trees can be changed by the therapist	High: therapy scenarios can be personalized for the each patient
Accessibility for the therapist	Low: the therapist can only spectate the scene from an additional monitor	Low: the therapist can only spectate the scene from an additional monitor and review logs after the session	Medium: the therapist can see and listen to the patient via audio and video feed. Special interface is provided for dialogue control	High: the therapist has multiple dashboards to track a patient's progress during the treatment
Affective state measurement	Low: only eye positions are tracked	High: psychological signals are tracked and analyzed to set difficulty levels	Low: patient's speech is recorded, but not analyzed	Medium: psychological signals are tracked, but affective state is not recognized
Scalability	Medium: architecture does not need extensive setup, however, it is based on a standalone solution	Medium: architecture does not need extensive setup, however, it is based on a standalone solution	Low: architecture requires a heavy setup	High: therapy sessions can be conducted online or via self-treatment

2.4 Affective computing in virtual environments

Affective state recognition, while using virtual environments, has been investigated by multiple researchers. Measuring and recognizing the emotional state of the user could lead to systems that can adapt to the user's emotional state.

The process of measuring such psychophysiological signals as ECG and EEG while the user is in a virtual reality environment was described by Marin-Morales *et al*

al. They constructed an experiment around 360° panoramas [61]. After viewing the panorama, sixty users immediately evaluated by it using self-assessment dummy for measuring arousal and valance. Their SVM-based model was capable to correctly classify arousal and valiance with 70% accuracy.

Mavridou *et al.* have used blood volume pulse (BVP) and electrocardiography (ECG) sensors to measure subjects' heart rate in virtual environments. The authors then used a support vector machine (SVM)-based model to recognize arousal and suggested that this kind of affective state recognition system would be useful for healthcare interventions [62].

The process of affective state recognition in dynamic virtual environments was described by Moghimi *et al.* While the experiments were not conducted in virtual reality (a VR headset was not used), it carefully describes every step of affective state recognition in virtual environments [63]. This paper describes GSR, EEG, and heart rate signal processing, normalization, and feature extraction techniques. KNN and SVM methods were used to classify binary emotional states of high and low arousal and anxiety with 98% accuracy. Finally, Moghimi *et al.* expanded their participant count to 103 and published an emotional state data set collected in virtual environments. However, they still merely used the game-virtual environment without VR headsets [64].

Christou *et al.* designed an affective state analysis system for virtual crowds [65]. This virtual reality environment was constructed with the idea that the user was surrounded either at a larger proximity or in close proximity when the virtual crowd avatars intruded the user's personal space. The authors used GSR and BVP sensors to measure the participants' reactions to the stimuli.

GSR, respiration, ECG, and EEG physiological signals were used by Wu *et al.* 19 participants drove a virtual Humvee in virtual reality during three military scenarios [66]. The first test was of a low threat, but, during the driving experience, the participants had to do a mental task, specifically, to name the color which popped up in the environment. The second test was the same as the first one, but in a high threat environment, and the third set was in the same high treat environment as the first one, but the participants were asked to do a Stroop test (the name of the color is written by using a different color) instead of color naming. The authors successfully recognized the participants' arousal.

Biofeedback based analysis for video games was described in the paper of Zafar *et al.* The authors constructed three video games and measured the participants' respiratory signals during the gameplay and stressful Stroop test tasks [67]. Biofeedback helped the players to recognize and adapt to their stressful behavior during the experiment.

Menezes *et al.* investigated possibilities of emotion recognition in virtual environments [68]. While the authors were not experimenting with virtual reality by themselves, they still argued that emotional databases, such as DEAP [69], where the stimuli were music video clips, can be used to train initial models to recognize the arousal and valiance of the user.

2.5 Recognition of stress and anxiety

Mental stress or anxiety detection is a highly popular topic of research [31][70]. Anxiety is recognized by using psychophysiological signals that are collected by using various sensors or variable devices. There are many signals which can indicate anxiety; usually, they include GSR, BVP, EEG, ECG, HR, speech and other observable measurements.

Gareth Walkom used biofeedback devices to measure psychophysiological signals [71]. Electrodermal activity and skin temperature were observed by using an *Affectiva Q Sensor* wristband. The results showed a significant increase in the psychophysiological signals.

Speech and GSR-based system was suggested by Kurnianwan *et al.* [72]. These signals were recorded while the experiment participants were solving light or heavy workload mental tasks. Additionally, the baseline of the relaxation period was recorded. Like in the previous studies, a higher accuracy was achieved by using signal fusion. Finally, 92% accuracy for three-class stress detection was achieved.

Gjoreski *et al.* [73][74] described a continuous stress detection system which used an *Empatica* wristband to physiological signals. Their experiments consisted of two parts – one in laboratory conditions, and the second one in the real life. During the experiments in the laboratory settings, the participants solved a stress-inducing task proposed by Dedovic *et al.* [75]. For real-life stress prediction, the participants were asked to manually note the hours that were stressful. The authors achieved 83% accuracy for the classification of ‘no stress’, ‘low stress’ and ‘high stress’ states.

Various factors can affect the reliability of mental stress (anxiety) recognition systems. One of the latest systematic reviews of stress recognition systems which used wearable sensors as the source identified the following main challenges [70]:

- Discrimination of similar emotions: the nature of human psychology can greatly impact the reliability of an anxiety recognition system. Some psychophysiological signals, such as the heart rate, can show up as a similar signal for different emotions. Thus, discrimination of the human emotions remains a challenge for all researchers.
- Unknown context and conditions: unless the experiment is conducted in perfect laboratory conditions, we must deal with non-restricted environments and stimuli. Even at the psychology clinic, we still cannot fully control such factors as the room temperature or audio-visual stimuli from the environment.
- Artifacts from movement: as the psychophysiological measurements are done through wearable sensors, the data can become full of artifacts. Even when the data is collected with the supervision of a professional, wearable devices can suddenly shift from their proper position. This is also a relevant issue for VRET studies due to the implied nature of 360° movement and vision.

- Subjective ground truth: in order to use classification methods, we have to label each point of the data and assign a specific value for it. However, as we do not know the ground truth, we have to trust the subject to correctly fill a subjective report, form or survey. Moreover, as these labels are subjective, they can change from person to person, even two subjects with identical psychophysiological signals can report different anxiety levels. Finally, the time to label the anxiety level (to fill a survey or questionnaire) also, can add major shifts in the labeling accuracy.

2.6 Psychophysiological signals for anxiety detection systems

The main factors influencing the design and the architecture of anxiety recognition systems are psychophysiological signals and, usually, non-invasive sensors to collect them. As the stimulus creates physiological responses, that leads to peripheral change and becomes an emotion.

There are several peripheral parameters that are used to recognize the affective state: galvanic skin response (GSR) or skin conductance (SKC), blood volume pulse (BVP), respiratory volume (RV), electromyogram signal (EMG), skin temperature (SKT) and heart rate (HR), which could be measured or estimated from the BVP signal sensor:

- Galvanic skin response (GSR) is a phenomenon which is manifested by the skin becoming a better conductor of electricity for a short period of time, when stimuli cause psychological arousal [76].
- Blood Volume Pulse (BVP) – the change of the blood volume measurement gives information about sympathetic activation. With the change of the diameter of blood vessels, changes of the blood volume and flow can be observed [77].
- Respiratory Volume (RV) – the calm state leads to slower and shorter breaths, while excitement and physical activity inflict deeper breaths. The stress state can be detected with more frequent breathing, while negative emotions result in irregular breaths [78].
- Electromyogram Signal (EMG) – emotional intensity leads to involuntary moderate contraction of muscles pushed by nervous energy [79].
- Skin Temperature (SKT) – changes to the skin temperature are induced by the vasodilation of peripheral blood vessels activated by the automatic nervous system [80].

- Heart rate (HR) – emotional experiences are often reflected through cardiovascular changes which are triggered to prepare for an emotional response [81].

Once the preferred psychophysiological signals have been picked, the sensors which will register these signals have to be chosen. Sensors can range from clinical ones to more user-friendly ones which do not interfere with the users' daily activities. The latest research often use *Empatica E4* or similar wristbands as they provide good quality of data and allow freedom for movement [82].

2.7 Psychophysiological feature extraction

The main psychophysiological signal features were reviewed by Picard *et al.* and are commonly analyzed in literature on emotion physiology [19].

These features included the mean absolute value of the raw signal, the standard deviation of the raw signal, the first difference of the mean absolute value for the raw and normalized signal, and the second difference of the mean absolute value for the raw and normalized signal.

Additionally, various extended statistical functions can be used to extract a variety of features that are used for the recognition of the affective state. Moreover, signal-specific features can be extracted from the GSR signal.

The BVP signal is revealed by the heartbeat-related feature set. Heartbeat can be calculated from BVP by using variational mode decomposition (**Figure 5**). VMD decomposes the signal into a number of modes of different center frequency, energy and bandwidth. Some of the decomposed modes then are dominantly modulated by the cardiac rhythm. These modes can be selected by using FFT-based analysis. Then, the maximum variation of modes is selected and used with the STFF algorithm to estimate the heart rate frequency [83].

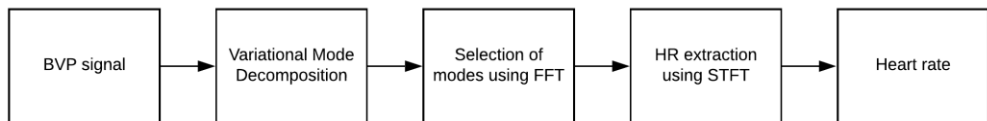


Figure 5. Process of decomposing BVP into heart rate

All of these features give a great variety of methods to extract the key changes of the signal that could indicate a change in the affective state. The whole feature set is shown in **Table 2**.

Table 2. Feature set of physiological signals

Signal type	Group	Feature
All	Statistical	Minimum $\min \{X_n\}$
		Maximum $\max \{X_n\}$
		Arithmetic mean $\frac{1}{N} \sum_{n=1}^N X_n$
		Variance $\frac{1}{N-1} \sum_{n=1}^N (X_n - AM)^2$
		Standard deviation (SD) $\sqrt{\frac{1}{N} \sum_{n=1}^N (X_n - AM)^2}$
		Median $(\frac{N+1}{2})^{th}$
		Kurtosis $\sum_{n=1}^N (X_n - AM) \frac{4}{(N-1)SD^4}$
		Skewness $\sum_{n=1}^N (X_n - AM) \frac{3}{(N-1)SD^3}$
	Expanded statistical [16]	Mean absolute deviation $\frac{1}{N} \sum_{n=1}^N X_n - AM $
		6 th moment $\frac{1}{N} \sum_{n=1}^N X_n^6$
		5 th moment $\frac{1}{N} \sum_{n=1}^N X_n^5$
		4 th moment $\frac{1}{N} \sum_{n=1}^N X_n^4$
		3 rd moment $\frac{1}{N} \sum_{n=1}^N X_n^3$
		Root mean square

		$\sqrt{\frac{1}{N} \sum_{n=1}^N X_n^2}$
	Differential [16]	$\frac{1}{N-1} \sum_{n=1}^N X_{n+1} - X_n $
		$\frac{\frac{1}{N-1} \sum_{n=1}^N X_{n+1} - X_n }{SD}$
		$\frac{1}{N-2} \sum_{n=1}^N X_{n+2} - X_n $
		$\frac{\frac{1}{N-2} \sum_{n=1}^N X_{n+2} - X_n }{SD}$
	Picard <i>et al.</i> [19]	Mean absolute of the raw signal
		$\frac{1}{N} \sum_{n=1}^N X_n$
		Standard deviation of the raw signal
		$\sqrt{\frac{1}{N} \sum_{n=1}^N (X_n - X_{mean})^2}$
		Mean absolute of the first difference of the raw signal
		$\frac{1}{N-1} \sum_{n=1}^{N-1} X_{n+1} - X_n $
		Mean absolute of the first difference of the normalized signal
		$\frac{1}{N-1} \sum_{n=1}^{N-1} \tilde{X}_{n+1} - \tilde{X}_n $
		Mean absolute of second difference of the raw signal
		$\frac{1}{N-2} \sum_{n=1}^{N-2} X_{n+2} - X_n $
		Mean absolute of the first difference of the normalized signal
		$\frac{1}{N-2} \sum_{n=1}^{N-2} \tilde{X}_{n+2} - \tilde{X}_n $
GSR	Peaks and amplitudes [84]	Number of GSR peaks
		$\sigma_k = \frac{1}{\lambda - 1} \sum_{n=1}^N \left[\left(m_{nk} - \frac{1}{\lambda} \sum_{n=1}^{\lambda} m_k \right)^2 \right]$
		Average GSR peak amplitude
		$ \bar{\sigma} = \frac{1}{N} \sum_{n=1}^N \sigma_n $

		Max SCR amplitude $\max\{\sigma_k\}$
BVP	Heart rate [85]	Beats per minute $BPM = \frac{60(t_n - t_1)}{n}$
		Interbeat interval $IBI = RR_t - RR_{t-1}$
		Root mean square of successive differences between adjacent R-R intervals $y_k = \frac{1}{W} \sum_{t=k-W}^k RR_t$ $\sqrt{\frac{1}{K} \sum_{k=1}^K y_n^2}$
		Standard deviation of successive differences between adjacent R-R intervals $y_k = \frac{1}{W} \sum_{t=k-W}^k RR_t$ $\sqrt{\frac{1}{K} \sum_{k=1}^K (y_k - y_{mean})^2}$
		Standard deviation if intervals between adjacent beats $\sqrt{\frac{1}{N} \sum_{t=1}^N (RR_t - RR_{tmean})^2}$
		Heart rate mean absolute deviation $\frac{1}{N} \sum_{n=1}^N HR_n - HR_{mean} $

2.8 Anxiety scales

As any classification system, mental anxiety recognition frameworks require a labeled dataset. Because of the difficult nature of human psychology, there is no ground truth for measuring the subject's anxiety. Therefore, researchers rely upon subjective scales and surveys reported by the subjects themselves. Here, we provide a list of widely used anxiety scales:

- Panic disorder severity scale [86] is a questionnaire created to measure the severity of the panic disorder. It is often used to determine the treatment outcome. This scale has seven items in the five-point scale,

in the range from 0 to 4. All scores are summed up and used for assessment. The final score ranges from 0 to 28.

- Anxiety change expectancy scale [87] is a 20 item questionnaire designed to determine and measure subjects' expectation in terms of anxiety treatment.
- Generalized Anxiety Disorder (GAD-7) scale [88] is a 7 item survey designed for rapid evaluation of anxiety and is used as a screening to measure severity indicators.
- Beck anxiety inventory [89] is a 21 item questionnaire with multi-choice self-report inventory. The scale has standard cutoffs:
 - 0–7: minimal anxiety
 - 8–15: mild anxiety
 - 16–25: moderate anxiety
 - 26–63: severe anxiety
- Subjective unit of distress scale (SUDS) [90] measures in scores from 0 to 100 (sometimes from 0 to 10), where 0 represents total relaxation and 100 represents the biggest anxiety attack in your life. The advantage of this scale is its simplicity and shortness that allows capturing anxiety in a specific moment.

2.9 Psychophysiological signal processing, feature extraction and classification methods for recognition of the emotional (affective) state

Emotional state recognition through psychophysiological signals requires a multi-step process and a variety of methods to be selected. The process starts from selecting the right signals for tracking and ends with feature extraction, affective state classification, and cross-validation.

The base of any emotional state recognition framework starts with signal selection. This is done by evaluating which psychophysiological signals are representing the emotional state in the most defining way as some emotional states can induce similar physiological signals, e.g., the heart rate. Additionally, the quality of sensors and the freedom of participant movement have to be taken into account. Medical grade sensors offer best quality and reliability, while wearable wireless sensors impact the subject to the least extent but tend to provide less quality for signal processing [91].

If signals are collected through wearable sensors, they tend to contain various artifacts and noise. Thus, it is suggested to perform signal smoothing by using a moving average or other similar filters [92]. Moreover, signals contain the dependency on the subject, e.g., the skin temperature of one person can always be

higher, or skin conductivity can be increased, which influences GSR signals. Therefore, signal normalization is applied before feature extraction [93].

The human body reacts to emotional stimuli in different rates. Therefore, signal windowing is used to find the length of the window which optimally reflects the emotional change. This is usually done by using a sliding window which cuts off the signal into separate samples for feature extraction. The most widely used window types are box, Tukey and Hamming [63].

Feature extraction is one of the most important parts of affective state recognition. The background which is now considered as the classical research was done by Picard *et al.* They demonstrated that statistical feature extraction can be used for a psychological signal like GSR, BVP, HR and respiratory volume to detect the state of arousal or valiance [94]. Moreover, emotional datasets, like DEAP by Koelstra *et al.*, collected data from 7 different physiological sensors, including EEG, extracted various statistical and frequency based features and matched them to the emotional states [69]. This was followed by hundreds of research papers applying different feature selection and classification methods.

When all features have been extracted, the stage of classification can begin. One of the most important parts in successful classification is the affective state labeling [95]. Most experiments use subjective questionnaires, like subjective units of distress (SUDS), BAI or emotional measurements, e.g., SAM manikins. In most cases, these surveys and questionnaires are converted to binary categories, e.g., high arousal/low arousal, and high valance/low valance for SAM manikins, or no/low/high anxiety for mental stress classification.

Finally, the constructed classification model has to be validated. There are several cross-validation techniques used by researchers: from the simple train-test split to 10-fold validation. However, while these popular methods provide optimistic results, the data samples of one participant can spill over to the general training set [96]. Therefore, it is better to additionally perform 10x10-fold cross validation, where we take 90% of the participant data for training and validate it against the remaining 10% by using the 10-fold method and averaging the results. Additionally, we leave one subject out of cross-validation can be performed where we take one subject out of training set and validate against it.

Virtual reality environments require that subjects should be able to move in 360 degrees and to have free hands if a VR controller is used. Thus, there are limitations with which sensors can be used and what signals can be tracked during the virtual reality session. As it can be seen in **Table 3**, measurement of the galvanic skin response (GSR) and heart rate (HR) is a widely used signal for the recognition of an affective state. Usually, sensors for these signals are relatively small and simple to setup and wear, thus, they are a good fit for virtual reality systems.

It is important for these systems to measure the subject's anxiety during the VR sessions, as it shows how the subject is reacting to the treatment. Still, most studies limit their signal analysis to the statistical relationships and do not use classification methods as this area is still new and requires collaboration between psychologists and software engineers.

Additionally, we can state that many studies use standalone and offline VR applications which do not offer a possibility of running VR sessions remotely or for the spectator (therapist/researcher) to change and adapt virtual environments in the real time. Most studies are conducted in laboratory-like conditions; therefore, these are not an important factor for the studies themselves. However, in order to popularize the usage of VR systems in real-life scenarios, more flexible, remote VR systems must be designed.

Table 3. Virtual reality systems with psychophysiological signal tracking

Author	VR usage scenario	Tracked physiological signals	VR architecture type	Observed affective states
Norrholm <i>et al.</i> [97]	PTSD	EMG, HR, GSR	VR offline (standalone, rail-motion)	Anxiety
Fusaro <i>et al.</i> [98]	Reactions to pain in others	GSR, HR	VR offline (standalone, no movement)	Reaction to painful/pleasurable stimuli
Detez <i>et al.</i> [99]	Gambling	HR	VR offline (standalone, no movement)	Arousal
Notzon <i>et al.</i> [100]	Spider phobia	GSR, HR	VR offline (standalone)	Anxiety
Breuninger <i>et al.</i> [101]	Agrophobia	GSR, HR	VR offline (standalone, movement with controller)	Anxiety
Kuriakose <i>et al.</i> [102]	Social interactions (autism)	BVP, GSR, skin temperature	VR offline (standalone, no movement)	Emotional state (angry, happy, neutral)
Gao <i>et al.</i> [103]	Moods, relaxation	EEG	VR offline (standalone, 360 images, no movement)	Attention, negative mood, positive mood
Czarnek <i>et al.</i> [104]	Training, rescue operations	ECG, HR	VR offline (standalone, movement with controller)	Reaction to rescue scenario

Esteves <i>et al.</i> [105]	Social phobia	GSR, HR	VR offline (standalone, no movement)	Anxiety (low/high)
-----------------------------	---------------	---------	--------------------------------------	--------------------

2.10 Conclusions of the Chapter

Anxiety disorders affect the daily life of more than a quarter of the world population. Usually, anxiety disorders are chronic conditions which require prolonged treatment. The ‘gold standard’ for anxiety disorder treatment is cognitive behavioral therapy. This kind of therapy uses multi-session exposure to the feared stimuli. This can be done either *in vitro* when the patient vividly imagines and describes the feared stimuli or *in vivo* when the feared stimuli are being faced in the controlled and safe real-world environment.

However, sometimes, it is difficult to organize *in vivo* sessions, as, for some types of anxiety, they require to be conducted in public places: airports for flight anxiety, traffic for driving anxiety, or an auditorium for public speaking anxiety. This is where VRET offers flexibility to reflect these environments in 3D space and help patients face their fears safely with the help of a therapist.

Multiple research studies have shown that VRET has limitations which are similar to the traditional therapy. VRET neither improves nor lowers the therapy efficiency when compared to control groups undergoing cognitive behavioral therapy without VR. Also, VRET suffers from the same drop-off rates of patients as the traditional sessions.

Moreover, VRET systems can be improved by anxiety recognition frameworks. The patient’s affective state can be tracked by using psychophysiological signals. Multiple researchers have shown that it is possible to process these signals and use feature extraction methods to prepare the collected data for classification. However, this kind of systems faces multiple challenges, such as artifact-heavy data from wearable meters to issues of setting a ‘ground truth’ for labeling the effective state when using subjective questionnaires.

Finally, from various papers analyzed in this chapter, we can concur that a virtual reality exposure therapy could be used as an effective tool for treating anxiety disorders. Thus, it is important to investigate various ways how VRET systems can be improved and made more widely available. Additionally, the following research should focus on enabling affective state recognition techniques in a VRET system for the possibility of personalizing VR scenarios.

3 MODELS AND METHODS FOR VRET WITH ANXIETY RECOGNITION¹

3.1 Analysis of the requirements for VRET system

Before designing a virtual reality exposure therapy system, the designers must consider several VR-related questions which are relevant to any virtual reality system design.

First, the content type of the virtual reality must be decided. The main options are either 360 degrees photos and videos or a fully 3D environment. The 360-degree images or videos could be cheaper to produce, although, when compared to fully 3D environments, they usually allow a broad yet smaller degree of possible interactions.

The creation of 3D content and environments can be a more expensive option. However, it depends on the type of the scene. The main benefit of creating a 3D environment is the freedom of movement and the possibility to interact with the objects within the environment.

The next decision is about the movement type. While 360-degree shots and videos are limited to jumps in space or rail-based camera movement, 3D environments have several options.

The first one is teleportation, where the user points to the surface in the 3D space, and, if the surface is valid, the virtual camera is moved to that spot. Another option is the movement by the control unit, when the user is moved to the camera direction by triggering a button on the controller or the headset itself. The final option is a rail-camera when the user is led through a pre-planned direction. Finally, the mix of all the movement types is possible.

After the content and movement type have been decided, the requirements of visual fidelity should be considered. Some virtual reality headsets connect directly to the personal computer, while others are based on dedicated or mobile devices. Such dedicated devices as *Oculus Quest* or *Oculus GO* are usually based on mobile phone-like specification with optimization for VR content processing, thus, they are limited on how detailed the 3D environment can be, as the drops in the frame rate during VRET sessions can cause the vertigo effect.

The real-life environment where the VRET system shall be used also has to be considered, as some VR headsets work only with additional sensors, and the empty space in the room must be dedicated to the VR setup. Additionally, some VR headsets are not mobile and have to be connected through cables.

If the designers choose the 3D content type, usually, a game development engine is picked to visualize and control the environment. The most popular picks among many researchers are either *Unity* or *Unreal*.

Finally, the hardware cost should be considered. Usually, mobile-based VR headsets are cheaper when compared to PC-dedicated VR hardware, e.g. *HTC Vive* or *Oculus Rift*.

¹ This chapter is based on author's work published in [107] and [106]

Once the requirements for the VR system have been defined, the system architecture design can be started. As any VR system can have multiple components, such as a VR headset, VR application, application to collect psychophysiological signals, network layer, etc., it is important to pick the right architecture type.

The requirements for an interactive VRET system can be derived from the considerations described in the previous paragraphs, and the focus is laid on the possibility for the therapist to adapt the selected virtual environment and scenario based on the patient's anxiety. This anxiety could be recognized based on the features extracted from the patient's psychological signals. While keeping these goals in mind, we can describe the key requirements for our VRET system design (**Table 4**).

Table 4. VRET system criterions and requirements

Criterion	Requirement
Interactions with virtual reality stimuli	The patient should be able to freely move in the virtual environment.
	Ability for the patient to interact with the key items (related to the feared stimuli) in the virtual environment.
	Virtual environments should be able to emulate sensations related to panic attacks (intensified breathing, audible heartbeats, etc.).
	The patient should be able to use the VRET system privately in a psychology clinic.
Multiple VRET patient's management	The therapist should be able to register new and edit already existing VRET patients.
	The VR-specific setting should be assignable for each patient separately.
	The therapist should be able to pick a specific VR treatment scenario for the specific patient.
	The therapist should be able to review the previous session log for the specific patient.
Virtual reality environment observability	The therapist should be able to see how the patient is behaving in the virtual environment and where the patient is looking.
	Interactions triggered by the patient should be synchronized with the therapist's view.
Virtual reality scenario adaptability	The therapist should have an option to adapt and change the virtual environment based on predefined options in order to lower or increase the VRET session's intensity.
	The therapist should be able to trigger the specific VR scenario flow.
	The therapist should have the possibility to change the reactions of 3D avatars and the expressed

	emotions.
	The therapist should be able to trigger visual and audio effects in the virtual environment.
Psychological signal collection	Psychological signals from wearable sensors should be collected and assigned to the specific patient and VRET session.
	Collected psychological signals should be cleaned from the noise and normalized.
Ability to recognize anxiety level	Relevant psychological features should be extracted from the recorded psychological signals.
	The patient's anxiety level should be recognized as based on the extracted features.
	The recognized anxiety level should be displayed for the therapist.

3.2 Virtual Reality as a Service (VRaaS) model

The VRET system was designed for the SaaS (system as a service) solution. Our goal was to develop a VRET system that could be accessible for psychology clinics anywhere in the world. Also, it was important for us to be able to scale the system up and to support multiple VRET sessions at the same time.

Therefore, the SaaS (System as a Service) model was adapted for Virtual reality therapy applications as the Virtual Reality as a Service (VRaaS) model.

The VRaaS model consists of five main subsystems: **(Figure 6)**:

1. **Affective state recognition** – this subsystem is designed for collecting psychophysiological signals, processing them, extracting various features that can be used in order to classify and recognize the current user's affective state.
 - 1.1. **Psychological signal capturing** – for capturing and storing psychophysiological signals collected via sensors from a patient that is using a VRaaS based system.
 - 1.2. **Psychological signal processing** – this component is designed to normalize and smoothen the collected signals and extract psychological features which could indicate the affective state change, such as the increased or decreased arousal.
 - 1.3. **Affective state classification** – based on the features extracted from psychological signals, the affective state is classified in order to pass it to the Virtual reality session management component.

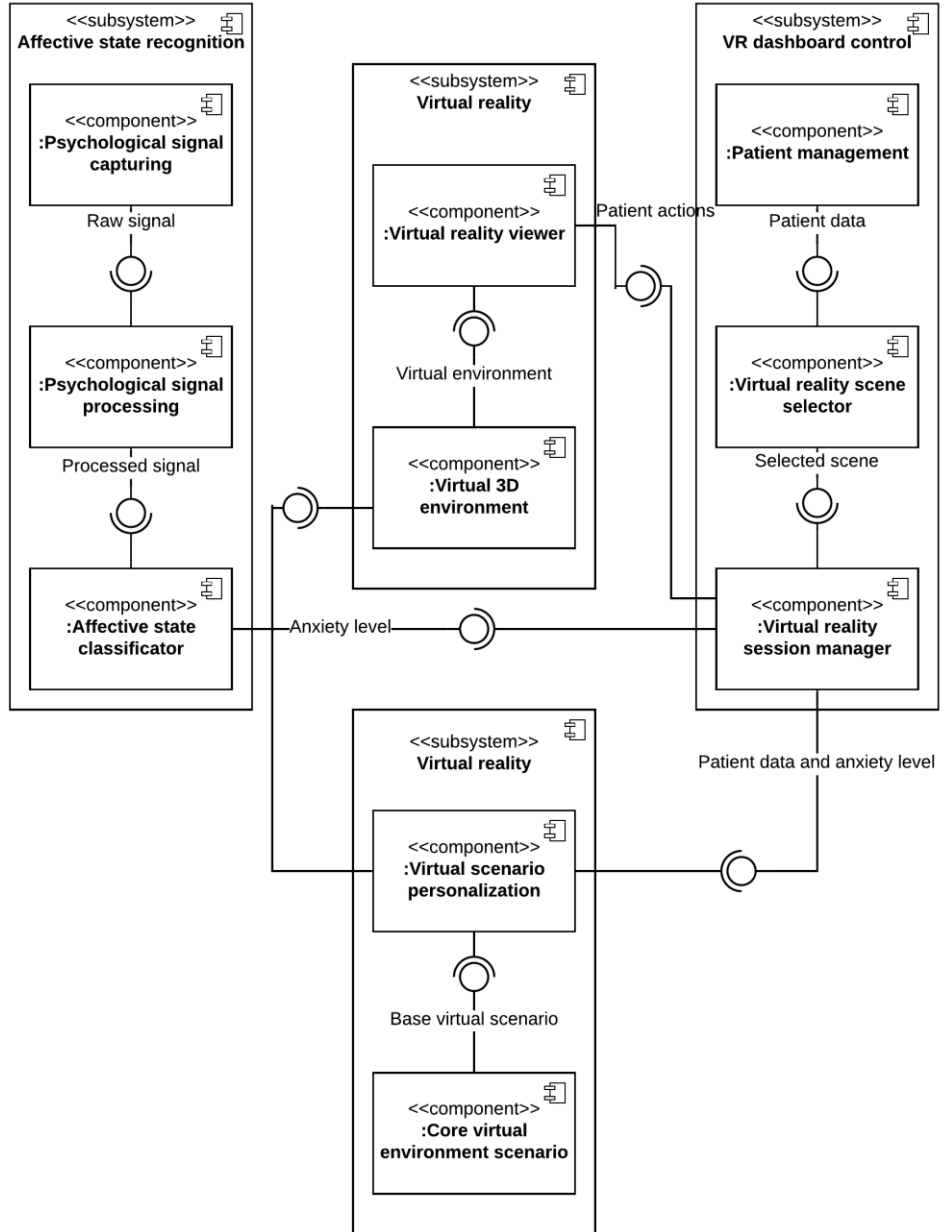


Figure 6. VRaaS model for Virtual Reality therapy system

2. **Virtual reality subsystem** – this subsystem is responsible for displaying the virtual reality content for the patient and managing the patient’s interactions in the virtual environment.

- 2.1. **Virtual reality viewer** – it is used as a tool to display a VR scene to the user and to capture the movement and interactions within the virtual environment.
 - 2.2. **Virtual 3D environment** – this component manages the state of the VR scene and the user's interactions. This component reacts to the VR scenario personalization changes that are initiated by the therapist and are translated to the virtual reality session management component.
3. **Virtual reality dashboard control subsystem** – it provides the possibility for the therapist to manage, select and control VRET sessions. Also, it provides a way for the therapist to interact with the scene and to personalize the therapy scenario.
 - 3.1. **Patient management component** – it helps the therapist to manage patients who are undergoing VRET therapy sessions and ensures the model scalability across multiple therapists and patients.
 - 3.2. **Virtual therapy scene selection** – this component is designed to manage the therapy scenes across multiple anxiety disorders and lets the therapist choose the appropriate scene for the patient.
 - 3.3. **Virtual reality session management** – this component displays the virtual environment for the therapist through the eyes of the patient and offers the possibility to interact with the virtual scene based on the virtual scenario personalization options.
4. **Virtual reality scenario flow subsystem** – it controls the flow of the virtual environment scenario based on the core workflow that can be changed by the therapist's interactions.
 - 4.1. **Virtual scenario personalization** – this component manages virtual scenario options that are available for the therapist to choose from in order to adapt the virtual environment, the avatar or this scene flow.
 - 4.2. **Core virtual environment scenario** – all the virtual reality therapy scenes are based on the core scenario which follows the baseline flow of the therapy session that can be adapted for the specific patient during the session by personalization options.

Based on the VRaaS model, we suggest an architecture for a web-based VRET system. The main components of our solution are the web service implemented by using the standard REST (Representational State Transfer) architecture, and the VRET dashboard application hosted on *Azure*.

The therapist is given access to their account in our system and can accept new patients, track the data of previous sessions and start a new session by using a VRET spectator. Each therapist is assigned a unique identifier.

This unique identifier is used in the *Unity* multiplayer service as a hosted game key, and it helps to synchronize the VR simulation between the Gear VR application used by the patient and the WebGL application used by the psychologist. This allows us to enable the therapist to view the patient's behavior during the therapy

session and to control the VRET session's intensity by changing the 3D environment.

The core parts of the deployment of our system can be seen in **Figure 7**. From the patient's side, we have chosen a mobile phone based VR headset application which communicates with the VRET service. This means that the patient only needs to start the VRET application, and the VR scenes, therapy scenarios and patient personal settings are synchronized based on the settings selected in the web dashboard, thus improving the simplicity of usage.

On the other hand, for the therapist, our system provides a web-based dashboard to control and manage the data of multiple patients. A web-based dashboard provides the system as a service approach for our VRET system, which means that VRET sessions can be managed from anywhere in the world and do not require any special hardware to be used; therefore, the therapist can use any laptop or personal computer with a web browser to work with the patients.

The communication between the patient's VR application and the therapist's VRET dashboard is arranged in the following way. Firstly, in order to accommodate a remote VR spectating option and possibility for the therapist to change the 3D environment and scenario in the real time, the socket-based communication between the WebGL based therapist application and the mobile VR application used by the patient was implemented. Both systems are built from the same source *Unity* engine game project and use standard multiplayer game design patterns. The WebGL application hosts 'the game' which, in our case, is the VRET 3D scene and scenario, while the patient's application joins this hosted game through the lobby system (the lobby is selected automatically based on the therapist's identification and does not require any input from the user).

The second communication layer is implemented through REST-based services. Both WebGL VRET application and the patient's mobile VR application are sending and receiving metadata about the ongoing VRET session. This metadata includes tracked psychophysiological signals, information about the selected VRET scenario, various settings of the current patient and the information about the required 3D assets to run the VRET scene itself.

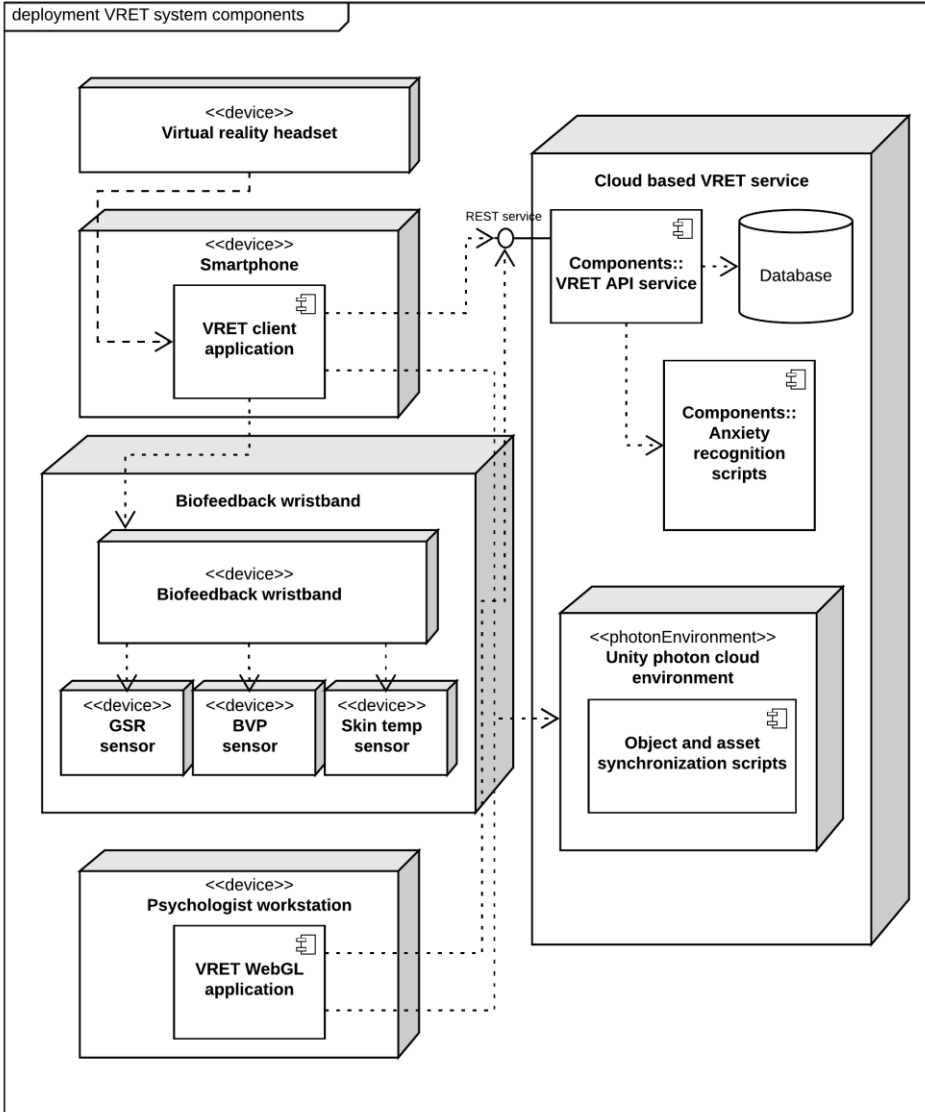


Figure 7. SaaS deployment for a virtual reality exposure therapy system

3.3 Proposed VRET system

In this section, we present the design of a VRET system with an anxiety recognition component which has been previously described in multiple research papers [106][107].

3.3.1 VR headset selection

For any VR solution, it is important to pick the correct virtual reality headset, as it is essential to the VR experience. We have reviewed widely used consumer grade VR devices. *Oculus Rift*, *HTC Vive* and *Samsung Gear VR* were considered as they are the most popular options. We have emphasized that it should be easy to set up the VR sensor, and it should require only minimal changes to the existing IT hardware. This was important as our target audience for the VRET system is psychology clinics. While HTC Vive and Oculus Rift can support more realistic 3D environments, they require a high-performance PC and an additional sensor setup in order to work.

The main goal of our web-based solution was to give access to the VRET scenarios through the internet and to make this access easy. This requirement led our decision towards *Gear VR*. The chosen headset can work without any sensors and with any mobile phone which supports VR applications. The use of mobile devices also sets low requirements for the psychologist's system. The service for the therapist with the embedded WebGL VR content can run on a wide variety of PCs and laptops which usually are already used in the workplace of the psychologist.

3.3.1.1 Psychologist's dashboard

The therapist's dashboard is an entrance point providing the capability to manage all the patients treated by the therapist. A VRET session can be started through the browser after a new patient has been registered. The patient's height is a key setting to provide good VR experience, as GearVR does not have a height sensor. The VR camera has to be configured to use this value as a reference point in order to avoid vertigo. The system's lightweight setup can be seen in **Figure 8**.



Figure 8. Minimal VRET system setup

The psychologist can start a VRET session by picking the type of fear it will focus on and the 3D scene scenario. The system will load a WebGL-based application which hosts a room in the *Unity Multiplayer* service, picked up and loaded by the GearVR application used by the patient. The patient's page, with the session history, is illustrated in **Figure 9**.

Our VRET service is focused around the interactions between the psychologist and the patient. However, the session must be started by the psychologist. When they log into the system, information about his or her patient can be entered. Previously conducted sessions are presented on the dashboard. Alternatively, the psychologist can register a new patient by filling out the correct information – the name, surname, and height. The psychologist and the patient can then start a VRET session. This is done by selecting an anxiety disorder and then selecting a specific VR scene. During the session, the psychologist can monitor the patient's performance and see additional information. The psychologist can start specific events inside the environment by creating situations which require the patient to use their social skills to comprehend them. After the session is over, the psychologist can review all the previous sessions.

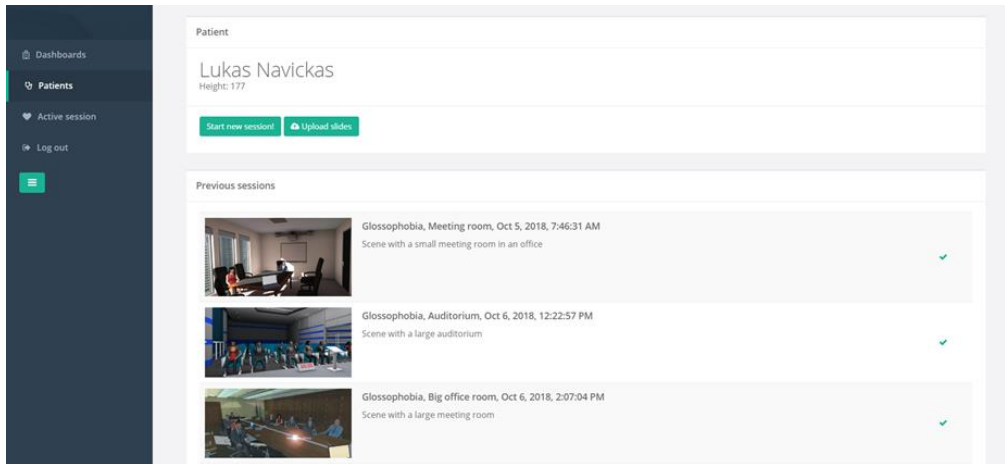


Figure 9. Patient overview page with a session history

3.3.1.2 VRET spectator's application

The *Unity3D* game engine was used to develop the spectator's application. This application relies on the WebGL technology and is served on the web browser. It communicates with the main components of the system by using the REST API. The spectator's application downloads the 3D data and loads the scene.

A unique identifier is used to create a game session with the same name in the *Unity Cloud Multiplayer* service. This hosted game includes a 3D scene with the simulation scenario. Only the Gear VR application which relates to the same identifier can join this VRET session. After a successful connection to the hosted game, the WebGL application starts to show the 3D environment seen through the VR headset in the web browser.

The psychologist can interact with the loaded 3D scene by using special control buttons on the WebGL application. Initiation of the interaction WebGL module fires the synchronization command in the *Unity Multiplayer* service, and the changes are transferred to the Gear VR application. Moreover, all the interactions made by the therapist are logged with the calls to the REST service. Therefore, all the collected records can be accessed and reviewed after the session is over.

3.3.1.3 VRET patient's application

The patient uses the Gear VR application created with the *Unity* game engine. At the initial stage of the setup, the user is prompted to enter the identifier provided by the therapist. This identifier is used by the GearVR application to find the patient on the database currently being used in the VRET session. By using this value with our REST API, all the collected data is assigned to the correct patient. Additionally, the unique identifier is used to join the correct game room in the *Unity* multiplayer service hosted by the psychologist on the WebGL application.

The selected VRET scenario scene is downloaded by the GearVR application from the asset repository. This is done by mapping the unique identifier to the VR scene. All the downloaded assets are cached within the mobile device and updated only when they have been changed. Finally, the VR application synchronizes the patient's behavior and movement in the virtual scene with the WebGL module running on the web browser.

3.3.2 Designing VR scenes for public speaking anxiety

Three VRET environments for the treatment of public speaking anxiety were designed and made available as part of our web service. During all the scenarios, the patients are put into a situation where they must deliver an oral presentation. The scenarios have different audience sizes and differing levels of social importance for the oral presentation.

3.3.2.1 VRET interaction

The therapist can interact with the VRET scenario by using the controls presented on the WebGL spectator's application. For public speaking scenarios, we implemented an emotional crowd control system. It enables differing 3D avatar reactions to be made to the speaker. Virtual listeners can demonstrate approval, disapproval, happiness and other emotions.

Moreover, the therapist can enable heartbeat or heavy breathing sound effects, which are often associated with panic attacks. Finally, an unexpected situation, e.g., a computer crash in the middle of the presentation can be simulated. This forces the patient to learn to adapt and stay calm.

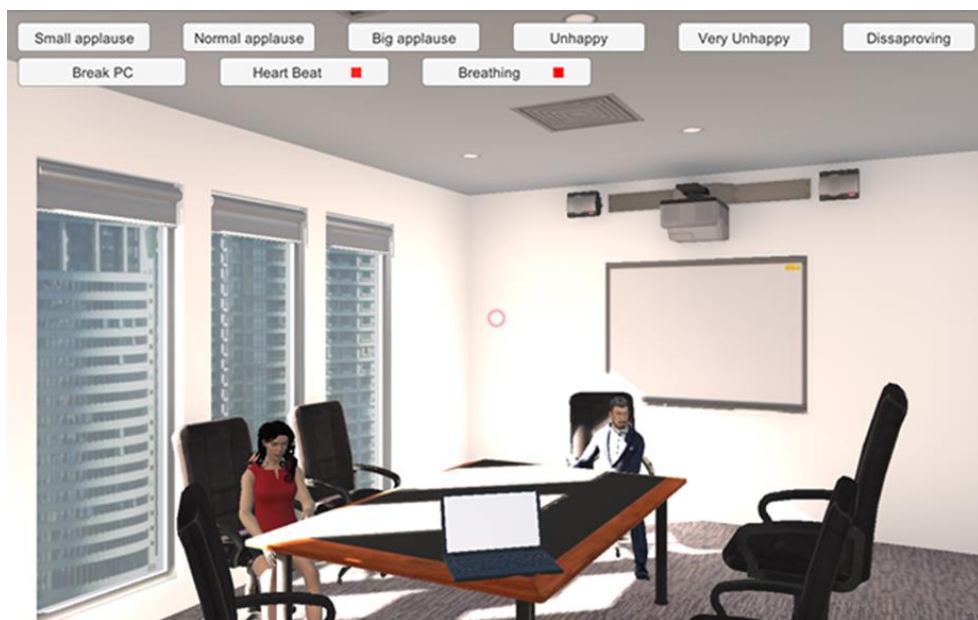


Figure 10. VRET control options

All available control options are illustrated in **Figure 10**

3.3.2.2 VR environments scenes

Three different VR scenes have been developed, from a presentation at a small office to a big meeting in the corporate environment, and, finally, a speech on the virtual stage in front of a big audience, with time pressure (**Figure 11**).

For VRET sessions, the patient must prepare a unique speech to be delivered with a slideshow presentation. These slides are uploaded to our service and are then integrated into VR environments. The patient can go back or forward in the slides by using the GearVR controller like a mouse clicker.

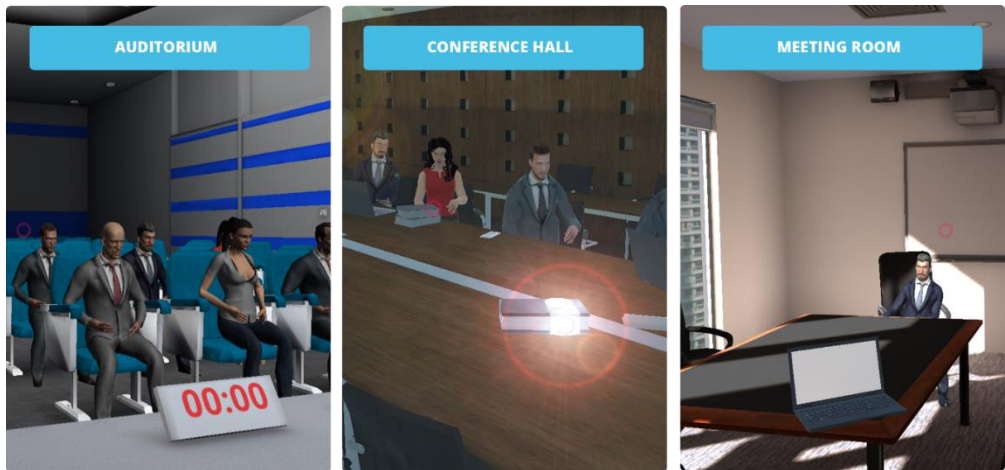


Figure 11. VRET environments created for the treatment of public speaking anxiety

Meeting room

The office meeting room is the smallest and least demanding environment. This scene is created for those who are getting started with the therapy and with the virtual reality in general. The patient is standing in front of two colleagues and must give them a presentation. In this scene, the psychologist can use different animations for the reaction of the avatars. For example, the therapist can initiate applause or disapproval reactions. The virtual PC that is demonstrating a presentation can also be turned off or restarted.

Finally, the psychologist can trigger such sound effects as the heartbeat or the sound of the patient's own breathing.

Conference room

The conference room is a much bigger scene than the office meeting room, and it contains a lot more complexity (lights, a projector, a white screen, etc.). The room contains an audience of virtual colleagues (3D avatars). The scenario flow of this environment is the same as in the previous meeting room scene. As in the first environment, the psychologist can interact with the virtual crowd in the room, set their behaviors, and add additional sound effects.

Auditorium

The auditorium is the biggest and most crowded scene of the three environments. In this scene, there are more than 20 virtual avatars, and there is a projector and an additional screen. There is also a timer which shows how much time the patient still has before the allotted presentation time is over. This scene is unique because it does not start in the auditorium: the patient must walk through the lobby and behind the scene before entering the stage.

3.3.3 Anxiety recognition in VRET system

To expand our developed VRET system with the possibility to detect anxiety during therapy sessions, we have implemented a mental stress prediction component which can be used during VRET sessions.

By using this component, firstly, we collect GSR, BVP and skin temperature signals. Secondly, these collected signals are processed by using signal denoising and normalization techniques. As these psychophysiological signals represent an autonomous response of the nervous system for the stress stimuli, physiological features can be extracted from these signals that indicate changes in the patient's anxiety.

An *Empatica E4* wristband was used to perform this task. This application collects signals from the wristband during the sessions and reports them to VRET REST API. This data is then assigned to a specific subject and a specific session. The consumer-grade device *Empatica E4* has a limited frequency for each signal BVP – 4Hz, GSR – 4Hz, and the skin temperature – 64Hz (**Table 5**).

Table 5. Signals (specification) collected in proposed VRET system

Signal type	Frequency	Metric
BVP	64 Hz	mmHg
GSR	4 Hz	μS
Skin temperature	4 Hz	°C

The collected signals and extracted multimodal features could then be organized into a dataset for anxiety recognition by adding the labels collected through the subject's psychological anxiety scales.

Finally, the organized dataset can be used to train an anxiety recognition model which can predict the patient's anxiety during virtual reality sessions and would help to the psychologist to correctly adapt the VR environment.

3.3.4 Software tools

The VRET system and the virtual 3D environments have been developed by using the *Unity* game engine. Most other works reviewed in Chapter 2.2 used either *Unity* or *Unreal* game development engines to implement their VRET applications. In this work, the *Unity* engine was chosen to implement VR scenes. This was motivated by the *Unity* libraries and SDK for a VR headset as the same code can be more easily reused for VR headsets produced by different manufacturers. Additionally, *Unity* provides full support for the WebGL runtime.

The therapist dashboard consisted from the WebGL application compiled with the *Unity* game engine. For the networking system between the WebGL application

running on the therapist's machine and VR application, the *Photon Unity 3D* networking framework was chosen. The *Photon* solution provides support for game rooms and lobbies that were used to connect the therapist's WebGL viewer and the VR application into the same virtual environment.

For the therapist's dashboard and system web application, the *.NET Web API* framework and *AngularJS* were chosen. This combination provided possibility to implement this service as a single page application, which reduces page reloading times and helps to implement modern UX (user experience) features on a web browser.

The physiological signals were analyzed by using *Python*. The most popular languages used in previous researches (Chapter 2.9) to analyze psychological signals are *Matlab* and *Python*. While *Matlab* has multiple plugins and extensions to analyze these signals, *Python* relies on a wide-open source community to create and support various packages for data and statistical analysis.

One of the most popular *Python* libraries for data handling and manipulation is *Pandas*. This library provides an advanced interface for data set transformation, filtering, and analysis. Thus, it was chosen for these tasks in our work.

For the numerical manipulation which was needed for signal filtering and the extraction of statistical features, we have chosen *Numpy* [108] and *scipy* libraries. They provide a standardized way to work with the numerical data and are used in most of the *Python* scientific libraries.

In order to analyze more specific features for GSR and BVP signals, the *BioSSPy* library was used for GSR signal analysis [109], and the *Heartpy* [110] library was used for the BVP signal. Both libraries were described in academic papers and presented at international conferences. Additionally, they were used by other authors in the field of affective computing.

To plot various graphs and signal visualizations, the *Matplotlib* library was used. *Python* has multiple libraries for plotting the data, but *Matplotlib* has a native connection with the *Pandas* library.

Finally, to train classification models, we have used the *scikit-learn* [111] library. This library is used by various researchers to train and validate their machine learning models, provides a library of machine learning methods and tools for result cross-validation.

3.4 Evaluation

To sum up, our proposed VRET system design and architecture offers several advantages:

- **Interactivity with the VRET scene:** by applying a multiplayer-based design in our VRET system, we have enabled the therapist to be not only a passive spectator, but also an active participant during a VRET session. It lets the therapist change 3D environments and therapy scenarios by adapting them to the individual patient's behavior during the session.

- **Online and remote usage:** the design and architecture of the system as a service for our VRET system enables the possibility for the therapist to access and use the system from any location via a web browser and does not require any additional hardware. The patient only has to have access to a mobile phone, a wearable sensor and a wireless VR headset.
- **Organized meta-data and psychophysiological signal tracking:** REST-based service and API enables the possibility to safely store metadata and psychophysiological signals that can be accessed afterwards and enables the possibility to integrate anxiety recognition into the VRET system.
- **Light system setup:** by choosing WebGL as a base for the therapist's VR application and a mobile phone for the patient, we have decreased the hardware requirements for the system itself. This means a simpler and less costly setup for psychology clinics.

3.5 Conclusions of the Chapter

This chapter reviews the requirements and design patterns for virtual reality systems, psychophysiological signal selection for anxiety recognition systems, and the architecture of the systems. Additionally, we describe the proposed system design for a VRET system with an integrated anxiety recognition component based on the analysis of a psychophysiological signal.

Any design of a virtual reality system starts with the definition of the system requirements, which leads to the selection of the correct virtual reality headset. Firstly, the format of the VR content must be selected between 3D environments and 360-degree photos or the video. Then, the movement type should be chosen: rail-camera, teleportation, or free movement triggered with controllers. Lastly, the quality of the required content has to be weighed against the system mobility, environment VR system costs, and hardware costs.

The core part of the anxiety recognition component is the psychophysiological signals and sensors that are used to classify the subject's anxiety levels. The widely used signals range from the galvanic skin response, blood volume pulse, electromyography and respiratory volume to electroencephalography. However, when selecting the signals and sensors, it is important to consider the virtual reality system itself. The sensors should not limit the movement of the subject, and, if possible, should not interfere with the VR headset; this is especially relevant for EEG sensors.

VRET systems offer a range of designs and architectures. One of the more popular VRET systems design patterns are standalone applications which are run on a single PC, and the therapist only acts as a spectator, or else they can be used by the

patient him/herself for self-treatment sessions. On the other hand, online VRET systems let the therapist conduct the session and observe the vital signals of the patient from remote locations. Finally, some VRET systems offer interactivity as the therapist can not only spectate the session but also change and adapt 3D environments according to the patient's behavior.

Our proposed VRET system with an integrated anxiety recognition component offers several advantages for the therapist. Firstly, a web-based dashboard allows adding and managing the data of new and already existing patients, to start and select the required VRET scenes which automatically start on the patient's VR application. Secondly, we have designed several environments for public speaking anxiety; these environments offer the possibility for the psychologist to change the crowd's reaction to the speech delivered by the patient, to interact with the environment by controlling the timer in the auditorium, the microphone echo, problems with the project delivery, etc., or by initiating audio-visual effects that are related to the panic symptoms, such as heavy heart beats, or extensive audible breathing.

Finally, while our designed VRET system has capabilities to include the anxiety level recognition component, it is important to clearly describe the method and the steps that are taken to achieve this within the VRET environment. This topic is widely discussed in Chapter 4.

4 VRET ANXIETY LEVEL CLASSIFICATION METHOD²

4.1 Anxiety level classification for VRET

In order to classify the patient's anxiety level during a virtual reality exposure therapy while using psychophysiological signals, several stages must be completed. As mentioned in the previous chapters, it is possible to record BVP, GSR and skin temperature signals during VRET sessions due to using a wearable biofeedback device. This data can later be used to train a model for anxiety level classification. The method to implement this can be deconstructed in these sections:

- **Signal collection** – during a VRET session, the patient's physiological signals are recorded by using a wristband device. These signals were synchronized and recorded three times during a session using an *Empatica E4* device (Section 3.3.3, **Table 5**). After each recording, the subject filled out a subjective anxiety scale so that the collected signals can be labeled with a certain anxiety measure (Section 4.6).
- **Signal normalization** – the collected signals have to be normalized in order to avoid subject dependency. The normalized value could thus be used for feature extraction.
- **Signal denoising** – due to the nature of a virtual reality system, the patient has a degree of freedom of movement during a VRET session. Thus, wearable sensors introduce an additional level of errors in the signal and must be denoised.
- **Signal windowing** – stress and anxiety triggers have different timing, and windows resonate through the autonomous nervous system. Therefore, the correct signal window size has to be determined for the most accurate results.
- **Feature extraction** – for anxiety classification, various psychophysiological features are extracted from the collected GSR, BVP and skin temperature signals.
- **Data labeling** – since there is no single ground true for anxiety measurement, psychological subjective scales are used to evaluate anxiety and to label the collected data samples.

² This chapter is based on the author's publication [136]

- **Anxiety level classification** – it is important to investigate various signals and dataset variations while training machine learning models for anxiety level recognition during VRET sessions. Signals can be used separately or as a multimodal input.

4.2 Signal normalization

When working with physiological signals, it is important to avoid subject dependency. One person's skin can be normally hotter than another one's, or sweating can be exceptionally high. Thus, the collected signals have to be normalized. It is important to note that some features require the range of (0,1) to work correctly, while other features can have negative values. Therefore, we have used two normalization methods. The first one is min-max normalization (2):

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (2)$$

here, X_n is the min-max normalized signal value, X is the value of the raw signal, $X_{\max}$ is the maximal value of the signal, and $X_{\min}$ is the minimal value of the signal.

The second one is zero, which means normalization (z-score) (3):

$$X_{zn} = \frac{X - X_{\text{mean}}}{X_{\text{std}}}, \quad (3)$$

here, X_{zn} is zero means the normalized signal, X is the value of the raw signal, X_{mean} is the mean of the signal, and X_{std} is the standard deviation of the signal.

By analyzing the normalized values, we can observe that the public speaking exercise in front of a therapist and during a virtual reality therapy session had a psychophysiological impact for most of the participants. The boosted GSR and skin temperature values usually imply increased anxiety. Additionally, as our participant group had various levels of public speaking anxiety, we can spot some outliers in **Figure 12**. They were not affected either by the live speech exercise during the session, or by the VRET session.

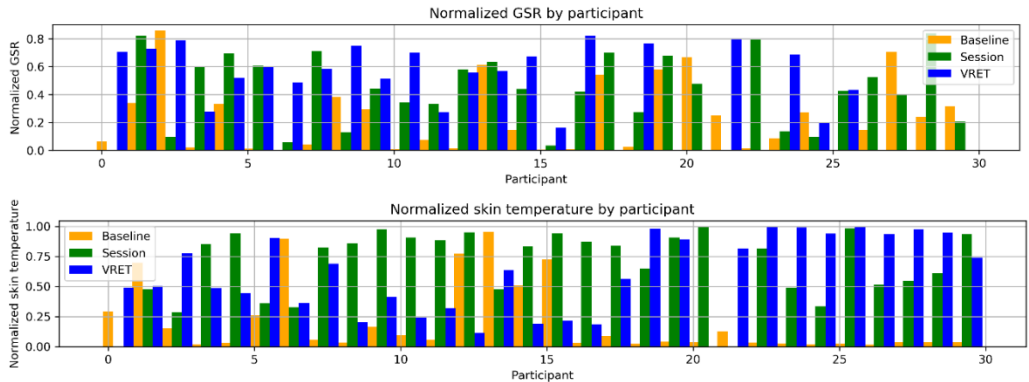


Figure 12. Normalized GSR and skin temperature session values by experiment participant

4.3 Signal denoising

The main initial challenge when processing physiological signals is the random noise. This was especially impactful to our data set as wristband tends to move and lose contact with the skin. We selected a moving average as one of the most common signal filters in order to reduce noise and artifacts while keeping the fast step response. This filter averages multiple input points to generate the output signal (1):

$$p[i] = \frac{1}{N} \sum_{j=0}^{N-1} s[i+j], \quad (1)$$

here, p is the output signal, s is the input signal, and N is the number of signal steps to put on the average.

We have applied the low pass 1Hz Butterworth filter in order to clean the collected data and to remove outliers. Moreover, a moving average filter was used to smoothen the signal. A comparison of the noisy raw signal and the processed signal can be seen in **Figure 13**.

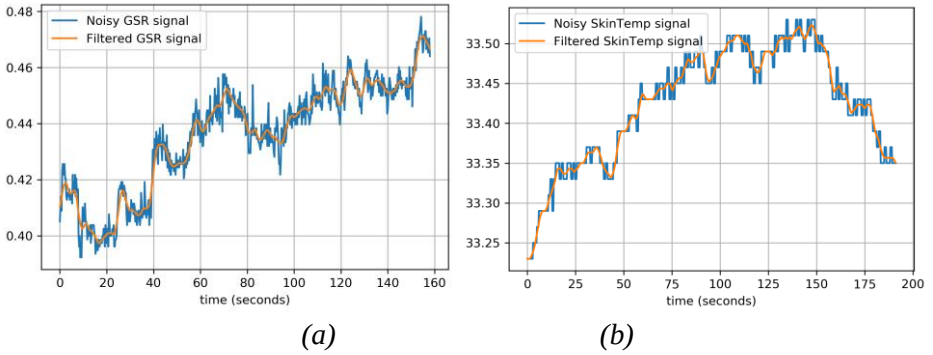


Figure 13. Raw and filtered biofeedback signals – a) GSR and b) skin temperature

4.4 Windowing

Different stimuli can have a varying emotional activation time. Therefore, it is important to select the correct signal window size in order to get the best prediction accuracy. We have decided to investigate which window size is best for VRET stimuli. A sliding window with a 1s interval was applied to all the collected signals with the following sizes in seconds: 3, 5, 8, 10, 13, 15, 18, 20, 23, 25, 28, 29 and 30.

4.5 Feature extraction

The most important part of creating a model which can recognize the anxiety level from physiological signals is the feature extraction. We have selected

the features which were suggested in literature, and grouped them by the signal type and origin.

Firstly, there are mathematical features that can be applied on all signals. Our statistical group contained the common statistical features that can be extracted with any math package.

Our extended statistical features were comprised of various statistical momentums and some less used statistical features. Our differential group included features with differential equations. Lastly, we dedicated a separate group for the psychological features researched by Picard *et al.* [94].

For the GSR signal, we have focused on finding the peaks and amplitudes of the skin conduction response (SCR). This was done by applying the Bartlett window and marking the SCR peaks between two consecutive zero-crossings. The amplitudes of the peaks were calculated by measuring the maximum value between the two points. Once all the peaks were detected, we calculated the number of peaks in the sliding window and the length of the maximum and mean amplitudes.

The BVP signal measures the volume of the blood that is passing with each pulse by measuring the volume in capillary beds on the skin. When processing BVP signals, it is important to extract heart rate features, inter-beat intervals and HRV features. All the extracted features are enumerated in Chapter 2.7 and **Table 6**.

Table 6. Features extracted from physiological signals

Feature name	GSR	BVP	Skin temperature
Minimum	X	X	X
Maximum	X	X	X
Arithmetic mean	X	X	X
Variance	X	X	X
Standard deviation	X	X	X
Kurtosis	X	X	X
Skewness	X	X	X
Mean absolute deviation	X	X	X
6 th moment	X	X	X
5 th moment	X	X	X
4 th moment	X	X	X
3 rd moment	X	X	X
Root mean square	X	X	X
1st difference	X	X	X
1st difference divided by SD	X	X	X
2 nd difference	X	X	X
2 nd difference divided by SD	X	X	X
Mean absolute of raw signal	X	X	X

Standard deviation of raw signal	X	X	X
Mean absolute of 1 st difference (raw)	X	X	X
Mean absolute of 1 st difference (normalized)	X	X	X
Mean absolute of 2 nd difference (normalized)	X	X	X
Mean absolute of 2 nd difference (raw)	X	X	X
Max peak amplitude	X		
Avg. peak amplitude	X		
Number of SCR peaks	X		
Beats per minute			X
Interbeat interval			X
Heart rate mean absolute deviation			X
SD between adjacent beats			X
Root mean of adjacent R-R intervals			X
Root mean of adjacent R-R intervals			X

4.6 Anxiety level class assignment

Each physiological signal time series for each participant had the associated SUDS evaluation. The minimal recorded SUDS score during our experiment was 10, while the maximum value was 80. The highest values of 90 and 100 were not recorded since they would indicate only extreme panic attacks.

In order to employ the classification, we needed to assign a class for each value. Thus, we have used some formulations from SUDS description and assigned four classes to low, mild, moderate and high anxiety (**Figure 14**).

This method is used by the majority authors who are working in the field of affective computing; however, it suffers from strong limitations. Usually, the subjective anxiety value is recorded only a few times per session (in the case of this research, it was recorded three times for one subject – baseline, public speaking and VRET evaluation). Yet, the signal window size can variate, so even for different signal window sizes, the same anxiety score is applied. This is the key limitation of any research based on emotion recognition which cannot be easily solved.

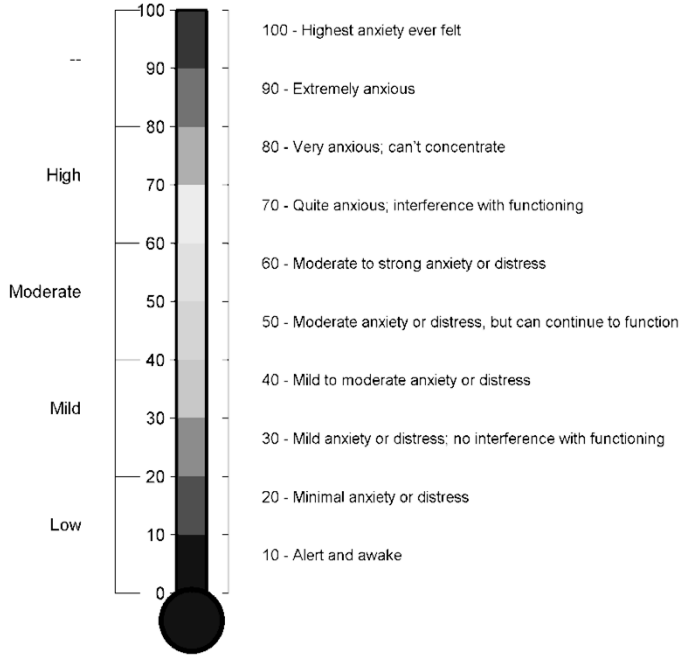


Figure 14. Anxiety level class assignment by SUDS value

4.7 VRET anxiety classification

After all the signals have been processed, all the selected features are extracted, and a feature set is formed, and all the data samples are labeled based on the SUDS scale; then, the anxiety classification models can be trained. Additionally, multiple classification models have to be trained in order to evaluate the influence of the signal window size and the signal type on the accuracy of the anxiety level classification. We have thus divided our analysis into multiple stages.

Firstly, we are analyzing different lengths of sliding windows. By this parameter, we are cutting the collected psychophysiological signals according to the selected window size with a one-second overlap. Secondly, we are preparing methods to investigate the importance of the signal feature for every signal separately and then for the signal fusion dataset. Lastly, we are evaluating the classification accuracy of GSR, BVP and skin temperature signals separately, and comparing them with the combined dataset.

All the classification models were trained by using a SVM classifier, as it was used by many previous studies for anxiety recognition. We have used the radial basis function (RBF) as the SVM kernel; penalty parameter C was set to 100, with C being class-balanced by automatically adjusting weights inversely proportional to class frequencies.

It is important to note that all the psychophysiological signals are collected during VRET sessions for public speaking anxiety that are led by a psychologist.

However, all the model training and data processing was done offline after all the VRET experiments had been finished.

4.8 Conclusions of the Chapter

This chapter reviews the methods that we have selected for psychophysiological signal processing, feature extraction and classification in order to recognize the subjects' anxiety during a VRET session.

For the stage of psychophysiological signals processing, we have performed signal normalization and signal smoothing as the data was collected from a wearable wristband *Empatica E4*, which causes additional noise. Normalization is done in two ways: based on min-max and zero means as some features require only positive values, while other features require both negative and positive values. Additionally, we have discussed the importance of the signal windowing as the stimuli of anxiety have some delay until anxiety can be recognized through physiological reactions.

Moreover, based on other studies in the field of affective state recognition, we have selected a psychophysiological feature set that can reflect changes in the subject's body during the state of increased anxiety. We have selected galvanic skin response (GSR), blood volume pulse (BVP) and skin temperature (SKT) as the key signals for our anxiety recognition component. They can be collected from a single wearable sensor (a wristband device) which does not limit the subject's mobility during a VRET session and is simple to set up.

Finally, we have selected an anxiety scale for our experiment. As there is no ground truth for measuring anxiety, we must rely on subjective anxiety questionnaires filled by participants. For this purpose, we have chosen subjective units of the distress scale as it offers a quick way for a subject to express his/her current level of anxiety.

5 EXPERIMENTS AND RESULTS ³

5.1 Participant demographics

A group of 30 subjects participated in our experiment. All the subjects gave and signed their informed consent for inclusion before they participated in the study. The study and protocol were approved by *Kaunas Regional Biomedical Research Ethics Committee* (reference number BE-2-60).

The group consisted of 13 females and 17 males aged 21–34. Most of the participants (76%) had a job. They were therefore familiar with the workplace public speaking scenarios (presentations during meetings in the office). The participants’ previous experience with virtual reality systems and video games was recorded. All the demographic data is shown in **Table 7**.

Table 7. Demographic information of the participants

Measure	Value
Age	M=26.85, SD=4.189
Sex	Female (13), Male (17)
Employment	Studies (7), works (23)
VR experience	Never (13), Rare (16), Frequent (4)
Video games experience	Never (10), Rare (9), Frequent (11)

5.2 Flow of the experiment

Each participant took part in a single therapy session. At the start of this session, the subject had to complete the initial demographic forms and the consent to participate in the experiment. Multiple psychological self-observed scales were presented to the patient, including: *Public Speaking Anxiety Scale* (PSAS) [112], *Personal Report of Communication Apprehension* (PRCA-24) [113] and *Speech Anxiety Thoughts Inventory* (SATI) [114]. Baseline galvanic skin response and skin temperature samples were collected from each subject. This data was later used as the baseline.

³ This chapter is based on the author’s experiments described in [136] and [107]

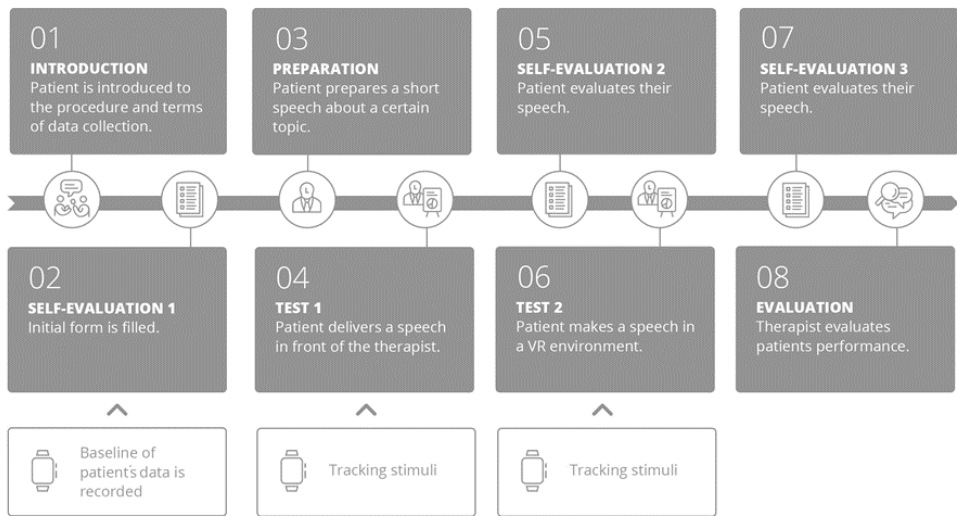


Figure 15. Flow of the experiment. VRET is conducted in Step 6

During the next stage of the session, the participant had a conversation with the psychologist and had 10 minutes to prepare a short speech on a suggested topic. The subject then delivered this prepared speech in front of the therapist, while biofeedback signals were being collected. The speech lasted up to 5 minutes. Finally, the participant discussed the speech with the psychologist and evaluated it. The participant also registered the Subjective Units of Distress Scale (SUDS) [90] and self-observed signs of discomfort (vertigo, muscle tenseness, etc.) [115].

The experimental session continued immediately with the VRET stage. The subject was given up to 10 minutes to rehearse a prepared speech. They then had to deliver the speech in front of the virtual auditorium while the psychologist controlled the 3D crowd and the environment. As in the previous steps, a biofeedback device was used to measure GSR and skin temperature following the VRET stimuli. This speech was limited to 5 minutes.

When the VRET session was finished, the participant evaluated the speech and filled out multiple scales: the Public Speaking Anxiety Scale (PSAS), the Personal Report of Communication Apprehension (PRCA-24), and the Speech Anxiety Thoughts Inventory (SATI), the Subjective Units of Distress Scale (SUDS), and the Igroup presence questionnaire (IPQ) [116]. This was done in order to compare the values before and after the VRET session.

Finally, the subject had a discussion with the psychologist about his or her experiences during the session. The flow of the conducted experiment is demonstrated in **Figure 15**. The total length of the session ranged from 45 to 90min.

5.3 VRET stimuli

There are multiple studies for emotion recognition while using videos, images or music as emotional stimuli [69][74][117][118]. Moreover, a recent survey

showed that most studies using machine learning to detect stress and anxiety are creating stress stimuli through mental tasks in the laboratory setup [70]. However, in our case, anxiety is induced through the VRET stimuli, which cannot be easily compared with video or images.

VRET sessions start with the participant waiting in the backstage of our virtual conference hall. After a short period of acclimation, the virtual character starts moving through the door to the virtual stage. Finally, the participant reaches the stage, is greeted by virtual listeners, and starts the presentation with the time limit. This initial period of the scenario acts as an emotional activator, and, from this point, we start to track the psychological signals of the subject. These stages of the virtual scenario are shown in **Figure 16**.

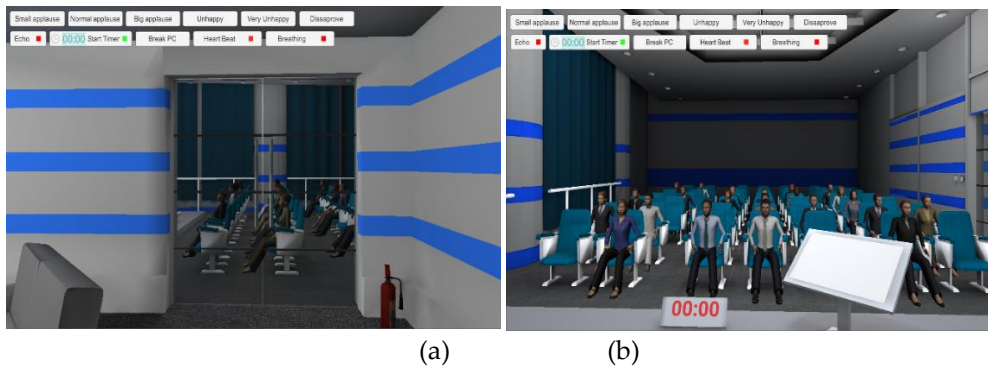


Figure 16. VRET public speaking environment: a) backstage; b) stage in the virtual conference hall

5.4 Psychophysiological signal collection through a wearable sensor

Our VRET system has been developed with the capacity to support wearable biofeedback devices. In this experiment, we have used an *Empatica E4* wristband [119]. The data from the skin temperature (ST) sensor was expressed in degrees on the Celsius ($^{\circ}\text{C}$) scale (sampled at 4 Hz), and the galvanic skin response in μS (sampled at 4 Hz).

During the entire experiment, the participants were using the wristband sensor *Empatica E4*. The sensors were connected with a computer by using Bluetooth connection. The acquired data included GSR, BVP and skin temperature throughout the session.

Three time series were recorded for each signal type and for each participant:

1. Baseline – recorded during the calm and relaxed state.
2. Speaking assignment – recorded during the speaking exercise in front of the psychologist.
3. VRET – recorded during the VRET public speaking scenario.

5.5 Anxiety level evaluation when using SUDS

Based on our experience from other VRET studies [120][121][122], in order to measure the participant's anxiety level throughout the experiment, we used a subjective unit of the distress scale (SUDS). This scale measures from 0 to 100 (sometimes from 0 to 10), where 0 represents total relaxation, and 100 represents the biggest anxiety attack in one's life. The advantage of this scale is its simplicity and shortness which allows capturing anxiety at a specific moment. All the values of SUDS are displayed in **Table 8**.

Table 8. Subjective Units of distress scale

Measure	Rating
Highest anxiety/distress ever felt	100
Extremely anxious/distressed	90
Very anxious/distressed; cannot concentrate	80
Quite anxious/distressed; interference with functioning	70
Moderate to strong anxiety or distress	60
Moderate anxiety or distress, but can continue to function	50
Mild to moderate anxiety or distress	40
Mild anxiety or distress; no interference with functioning	30
Minimal anxiety or distress	20
Alert and awake	10

5.6 Results

5.6.1 Collected psychophysiological signals

The subjects rated their anxiety levels during VRET sessions much higher than the baseline: 37.9 SUD against the baseline of 28.3 SUD. This indicates that the proposed VRET system and virtual environments provided stimuli for anxiety. All the reported SUD values are displayed in **Figure 17**.

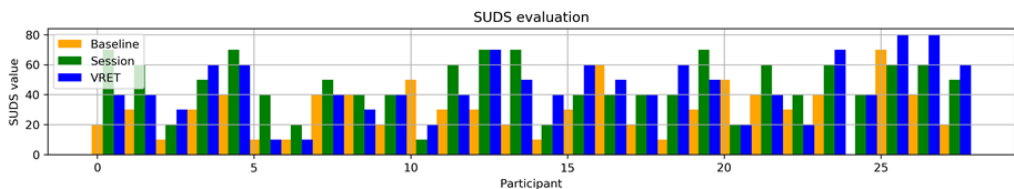


Figure 17. Values reported on the subjective units of the distress scale

We have also compared SUDS values of the subjects who agreed or strongly agreed with the statement “While preparing for giving a speech, I feel tense and nervous,” which has been included in the PSAS scale (a mean of 24.2 SUD baseline and 45.7 VRET session SUD, higher than the mean SUD). Comparison graphs are presented in **Figure 18**.

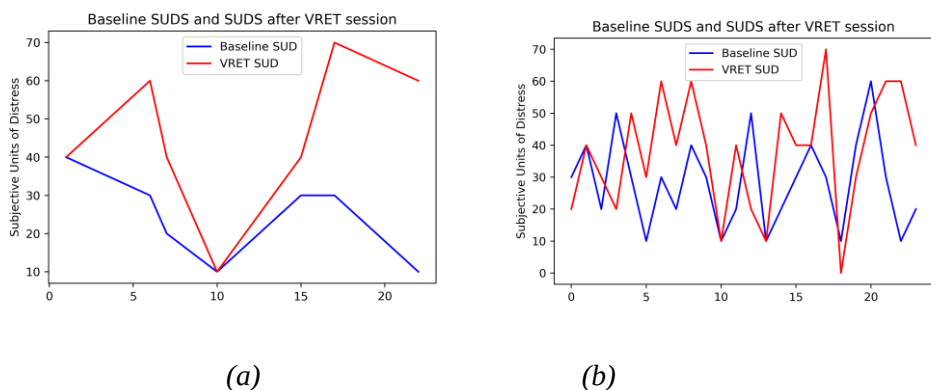


Figure 18. Subjective units of distress evaluation: a) those who agreed or strongly with the statement “While preparing for giving a speech, I feel tense and nervous”; b) all the remaining subjects

Additionally, we have compared the differences in the reported SUDS values based on the participants’ previous experience with VR applications and video games. We have noticed that gamers and those who tried VR before the therapy session reported higher SUDS values after the VRET session. This could indicate a deeper immersion within the virtual environments because of more experience with the virtual worlds (games). However, a dedicated study is required to investigate this effect further (Figure 19).

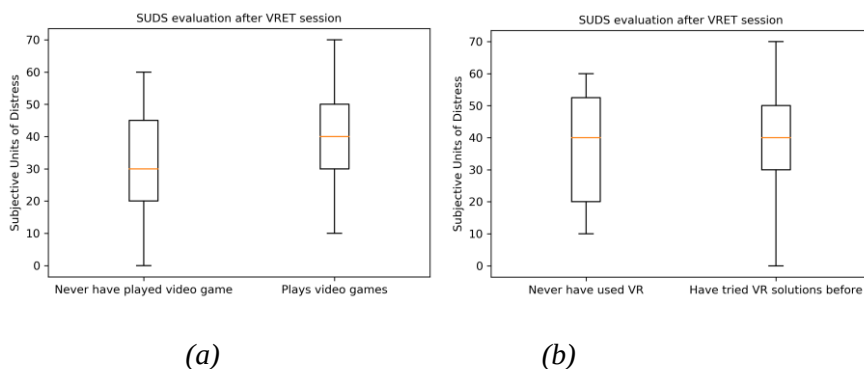


Figure 19. Differences in the SUDS values reported after a VRET session: a) based on the previous gaming experience; b) based on the previous VR experience

Some negative effects of virtual reality systems were noticed. This was achieved by analyzing subjective symptoms of discomfort reported by the participants. As expected, one of the negative reactions to the VR environments is th

e vertigo effect which was indicated by nausea (1 subject) and loss of balance (6 subjects). Moreover, wearing a VR headset increased sweating (6 subjects reported it during the speech exercise and 9 subjects were affected during the VRET session). The rest of the observed symptoms which were associated with the increased anxiety were similar in both cases. Self-observed symptoms of all the subjects are shown in **Table 9**.

Table 9. Subjective signs of discomfort

Measure	During public speaking	Public speaking during a VRET session
Increased sweating	26.09%	39.13%
Dizziness	8.70%	13.04%
Increased heart rate	78.26%	69.57%
Shivers	21.74%	21.74%
Uneasiness, fear	69.57%	60.87%
Heavy chest	8.70%	8.70%
Muscle tension	30.43%	43.48%
Loss of balance	8.70%	26.09%
Nausea	0.00%	4.35%

The loss of balance is also a side effect of the virtual reality systems and could limit the adaptation and acceptance of such systems. However, all except for one of the patients who experienced this symptom while using our virtual reality system had never used VR headsets before. Therefore, for the patients who have no previous VR experience, we would suggest offering a neutral introductory virtual reality session before any treatment sessions. This could possibly lead to better acclimatization in VR environments and to minimizing the loss of balance symptoms during VRET sessions.

Three series of the collected data were analyzed in order to evaluate the effects of the VRET system. The first series included the values recorded during the beginning of the experiment (the baseline), the second series included the values recorded throughout the speech exercise in front of the therapist, and the third series contained the values recorded during the VRET session.

By comparing the raw mean values of the signals from all the participants, we can see a significant increase of the values of GSR and skin temperature measurements. GSR signals were significantly more variable during the VRET session, which indicates an increased reaction to the stronger stimuli of anxiety (**Figure 20**). This increase demonstrates higher participant involvement in the public speaking exercise during the VR simulation than during the live speech.

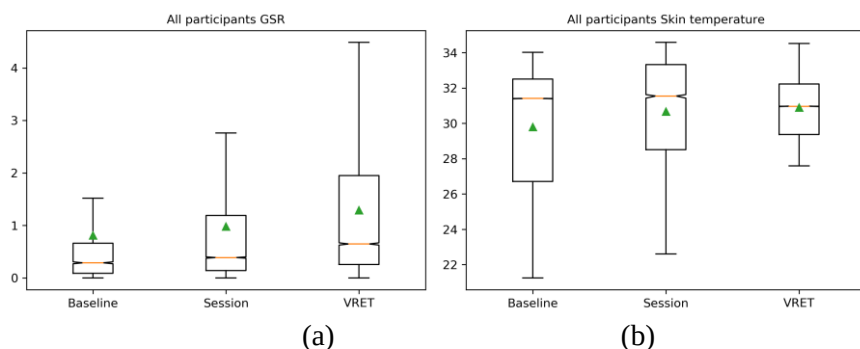


Figure 20. Normalized GSR and skin temperature session values by participant of experiment

Finally, by looking at the values of the extracted features, we can state that VRET speaking scenarios created a stimulus which was equivalent to or even stronger than a speech that was delivered in front of the therapist (**Table 10**). The values increased during the live and VRET session when we compared these features to the baseline value. Therefore, a web-based VRET service could be used as a helpful exposure therapy tool for treating or even self-treating public speaking anxiety disorders. By providing patients with the ability to emulate and prepare for their daily life challenges, we could enable them to adapt their reactions with or without the help of a therapist.

Table 10. Elevated GSR and skin temperature feature values

Feature	Baseline GSR / Skin temp.	Session GSR / Skin temp.	VRET GSR / Skin temp.
Raw signal mean absolute	1.152 / 29.74	1.257 / 30.72	1.372 / 31.21
Raw signal standard deviation	0.07 / 0.04	0.22 / 0.023	0.106 / 0.019
Raw signal absolute mean of 1 st difference	0.0044 / 0.0015	0.0121 / 0.011	0.0065 / 0.0011
Normalized signal absolute mean of 1 st difference	0.056 / 0.058	0.036 / 0.018	0.021 / 0.019
Raw signal absolute mean of 2 nd difference	0.0008 / 0.0003	0.0019 / 0.00028	0.0014 / 0.0003
Normalized signal absolute mean of 2 nd difference	0.0116 / 0.0149	0.0077 / 0.0054	0.0049 / 0.0057

5.6.2 Class distribution

The class distribution in our research mostly focused on mild and moderate anxiety (**Figure 21**). Uneven class distribution can lead to model bias to a specific class, as, during the training of the classification model, the trained model may tend to pick a class with the majority of samples in the training set.

One of the approaches towards avoiding this prediction bias for imbalanced data sets is to change the misclassification penalty per class [123]. In the SVM method this is referred as a class-weighted SVM. Normally, for SVM, we have to set only a single value for regularization parameter C. When class-weighted SVM is being used, the misclassification penalty for the minority class is picked to be higher than for the majority class.

Another approach is to either undersample or oversample the training data set. When random oversampling is used, the training data is supplemented with multiple copies of samples from the minority class. Meanwhile, random undersampling is balancing the data set by removing duplicate samples from majority class.

Moreover, for more advanced oversampling, SMOTE (Synthetic Minority Over-sampling technique) can be used. Synthetic data samples for the minority data class is created based on the feature space. This can be illustrated as taking the existing sample and k nearest neighbor and multiplying the vector between them with random number x from 0 to 1.

In our work, we have used a class-weighted SVM to adjust misclassification penalties. Then, we have measured prediction accuracies for each class separately. Finally, we demonstrate that our models do not have a strong pull to one particular class by plotting normalized confusion matrices from **Figure 27** to **Figure 30**.

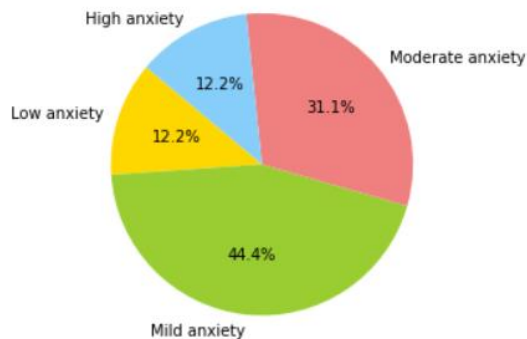


Figure 21. Anxiety class distribution

5.6.3 Validation methods

We have decided to use the accuracy metric for model-by-model comparison, as the majority of articles on anxiety prediction publish their results by accuracy. For cross-validation, we have used 10x10-fold validation, as, in this way, we are first dividing participants into 10-fold segments and using data of 90% of the participants

for training (27 participants), and the remaining ones for testing (3 participants); we are repeating this process 10 times across 10 folds. Thus, there is no data sample spillover from the training set to the test set [72][85].

Additionally, we have tested our models with the randomly assigned class to ensure that they are not biased, and all the discussed models had 25–26% accuracy, which is expected from a four-class classification model with random labels.

5.6.4 Signal window size evaluation

The windows size can severely impact the results of a classification. Therefore, we have trained all our models by using 3, 5, 8, 10, 13, 15, 18, 20, 23, 25, 28, 29 and 30s window length.

For GSR and BVP signals, the window size had a significant impact to the results, while the skin temperature was less impacted by the window size. For all the signals, the accuracy steadily increased as the window size was increased. The peak was reached at around 18–23 seconds.

The visual comparison regarding accuracy dependency on the window size is shown in **Figure 22**.

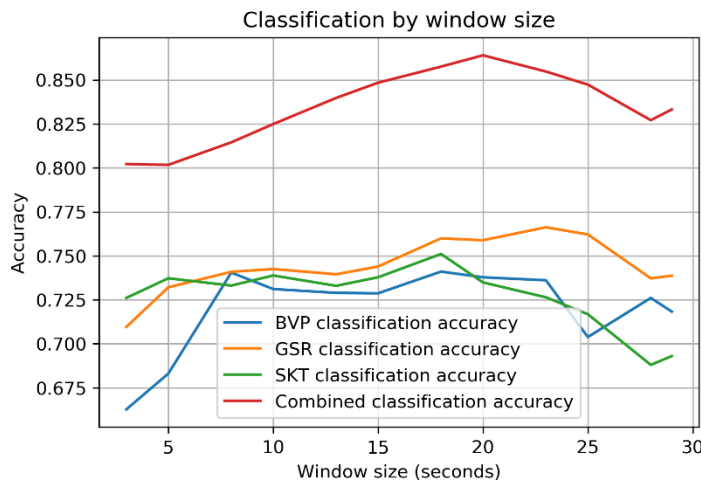


Figure 22. Anxiety level prediction by using blood volume pulse (BVP), galvanic skin response (GSR), and skin temperature: accuracy dependency upon signal window length

Anxiety level prediction by using BVP, GSR and skin temperature: accuracy dependency upon the signal window length.

Based on our findings, we can conclude that it is best to use the signal window size of around 20 seconds. For a smaller window length, the accuracy is lower, however, a long window size has a downside of the slower update rate as it would take around 20s to update the affective state change. A detailed comparison of results is listed in **Table 11**.

Table 11. Anxiety level prediction accuracy by signal type and window size (best results are shown in bold)

Window size, s	Accuracy using GSR	Accuracy using BVP	Accuracy using skin temperature
3	70.9%	66.2%	72.6%
5	73.2%	68.3%	73.7%
8	74.3%	74.0%	73.3%
10	74.3%	73.1%	73.8%
13	74.0%	73.8%	73.2%
15	74.4%	73.7%	73.7%
18	75.8%	74.1%	75.1%
20	75.9%	73.8%	73.3%
23	76.6%	73.0%	72.6%
25	76.2%	70.2%	71.7%
28	73.7%	72.6%	68.8%
30	73.8%	71.8%	69.1%

5.6.5 Feature evaluation

We have applied the random forest (RF) for feature selection. The method can expose feature importance for a classification task by using the Gini index [124]. This index shows a correlation between the signal feature and the anxiety label. The stronger relationship is indicated by a higher index value.

According to the evaluation of the Gini index, the most relevant BVP features include both statistical (Picard *et al.*) and extracted heart rate features (**Figure 23**) – *mean* (mean value of the normalized signal), *rms* (square root of the mean square signal value), *median*, *rawabsdiff1* (mean absolute of first difference of the raw signal), *rawabsdiff2* (mean absolute of second difference of the raw signal), *rawstd* (standard deviation of the raw signal), *min* (minimal value of the normalized signal), *mad* (mean absolute deviation of the signal), *maxHR* (maximum heart rate), *max* (maximal value of the normalized signal).

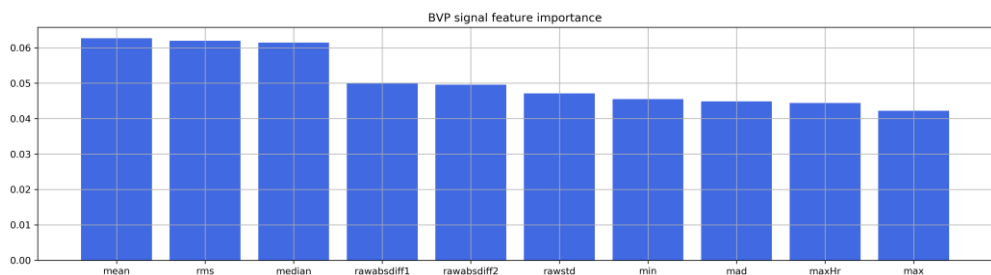


Figure 23. Feature importance BVP (ten most relevant features are shown)

GSR statistical values come out as one of the strongest features for this type of signal. Though, we can see that some of classical features from the research of Picard *et al.* were also included (**Figure 24**) – *min* (minimal value of the normalized signal), *max* (maximal value of the normalized signal), *median* (median value of the normalized signal), *rms* (square root of the mean square signal value), *rawmean* (mean value of the raw signal), *mean* (mean value of the normalized signal), *maxamp* (maximal amplitude of the signal), *rawabsdiff1* (mean absolute of first difference of the raw signal), *rawabsdiff2* (mean absolute of second difference of the raw signal), *absdiff1* (mean absolute of first difference of the normalized signal).

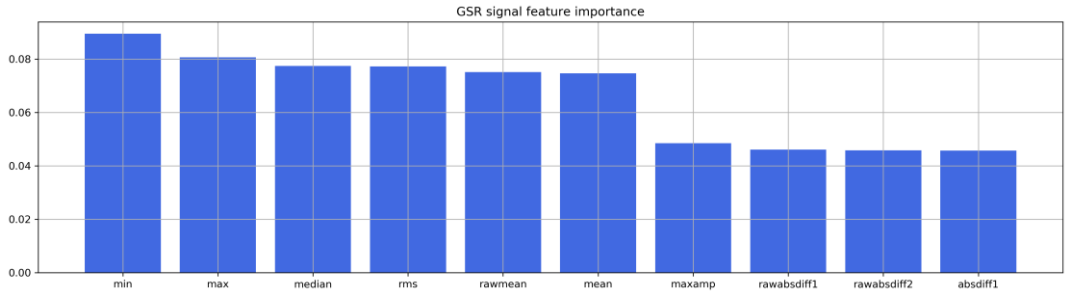


Figure 24. Feature importance GSR (ten most relevant features are shown)

As the skin temperature is similar and can be impacted by environmental factors, and as our experiment took place in a room with the temperature control, it is interesting to note that the raw mean value was among the best features (**Figure 25**) – *rawmean* (mean value of the raw signal), *rms* (square root of the mean square signal value), *mean* (mean value of the normalized signal), *max* (maximal value of the normalized signal), *median* (median value of the normalized signal), *min* (minimal value of the normalized signal), *absdiff2* (mean absolute of second difference of the normalized signal), *absdiff1* (mean absolute of first difference of the normalized signal), *zero* (zero crossing), *std* (standard deviation). However, normalized values were not far behind.

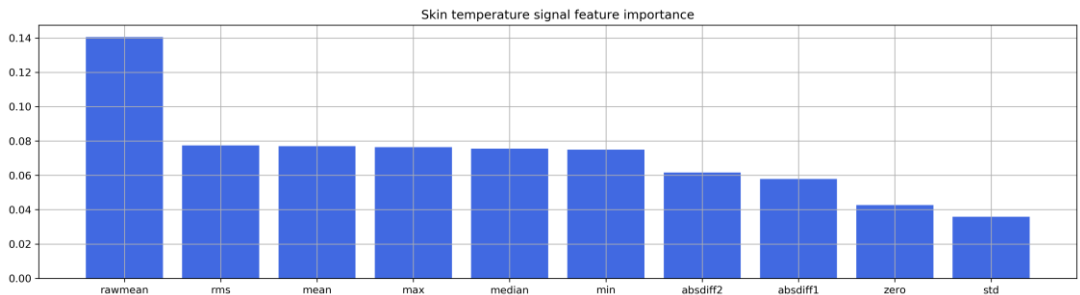


Figure 25. Feature importance skin temperature (ten most relevant features are shown)

In the case of signal fusion, we can state that strongest features were extracted from all signals. Interestingly, the method selected 6 features from skin

temperature, 3 features from BVP, and one feature from GSR. Ten most impactful features of the fused signal (with the highest Gini index values) are displayed in **Figure 26** – *skintemp_rawmean* (mean value of the skin temperature signal), *gsr_rawmean* (mean value of the GSR signal), *bvp_mean* (mean value of the BVP signal), *bvp_median* (median value of the BVP signal), *skintemp_max* (maximal value of the normalized skin temperature signal), *bvp_rms* (square root of the mean square BVP signal value), *skintemp_mean* (mean value of the normalized skin temperature signal), *skintemp_median* (median value of the normalized skin temperature signal), *skintemp_min* (minimal value of the normalized skin temperature signal), *skintemp_rms* (square root of the mean square normalized skin temperature signal value).

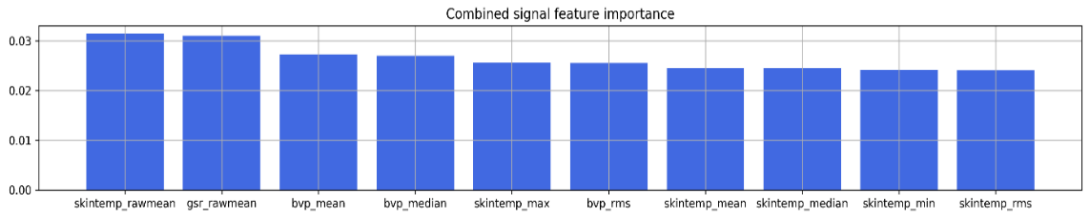


Figure 26. Feature importance signal fusion (ten most relevant features are shown)

5.6.6 Psychophysiological signal evaluation and signal fusion

When evaluating our final models, we are focusing on confusion matrixes. In multi-class classifications, sometimes classifiers tend to over-fit and can be biased to the particular class. However, in our case, all the four classes have strong prediction accuracies, and the classifier is not leaning on one majority class.

The classification using only the features from the GSR signal had 84.9% accuracy, which was the lowest prediction accuracy of all the signals. Usually, GSR is a good estimator for anxiety-based recognition [76]. However, we used a wristband-type sensor instead of a finger one, which is more prone to artifacts due to the movements of the subject (**Figure 27**).

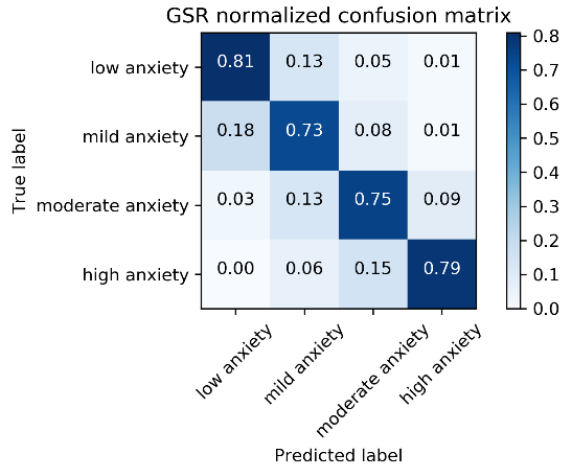


Figure 27. GSR signal classification normalized confusion matrix

BVP as the one of the most prominent indicators for stress recognition provided the best accuracy for single signal modality (**Figure 28**). From BVP, we can extract such features as the heart rate, heart rate variability, RR intervals and other values. However, sometimes, two identical BVP signals can reflect different emotional states [125].

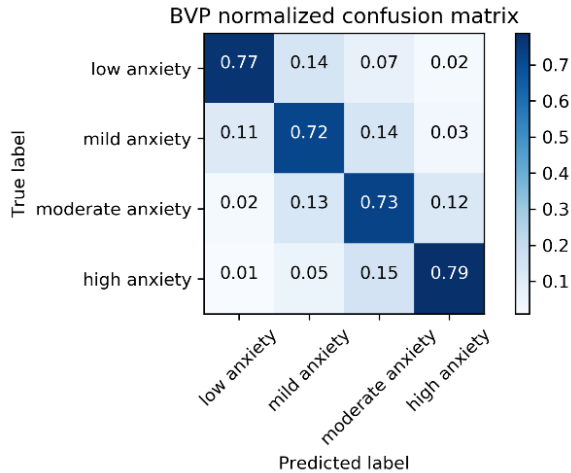


Figure 28. BVP signal classification normalized confusion matrix

Variations in the skin temperature are signs for the blood flow change in the skin. Other researchers showed that the change of arousal also impacts the skin temperature [126]. In our experiment, the features of the skin temperature were capable to detect the biggest swing of anxiety (low and high), while the recognition of the middle states was worse (**Figure 29**).

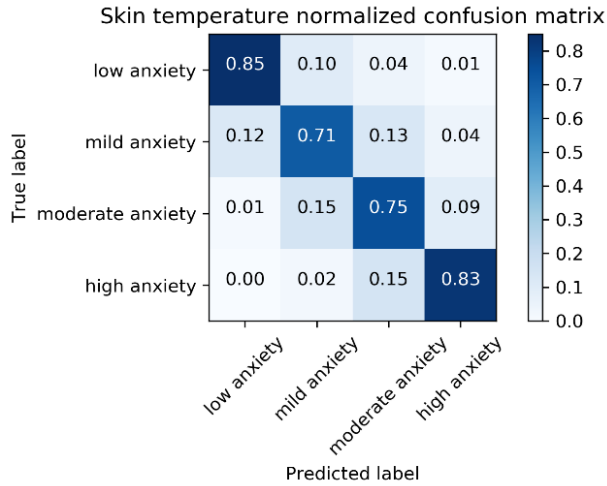


Figure 29. Skin temperature signal classification normalized confusion matrix

As in many previous studies, we have tried the signal fusion approach for all the modalities. There are several methods how signal fusion can be implemented. The most notable ones are early and late fusion approaches. During the early fusion, the signals from all modalities are mixed together and go through the feature selection stage as a combined dataset. Alternatively, during the process of late fusion, the signals from all the modalities are kept separate, and signals are fused only at the classification stage.

We have achieved the results where the model that was using signal fusion outperformed models that used standalone GSR, BVP and skin temperature signals. In our case, the method of early fusion has performed slightly better than the late fusion, by 86.3% and 83.2%, respectively. The accuracy of the early fusion was on point for most of the classes, although it had a small tendency to identify moderate anxiety as high (**Figure 30**).

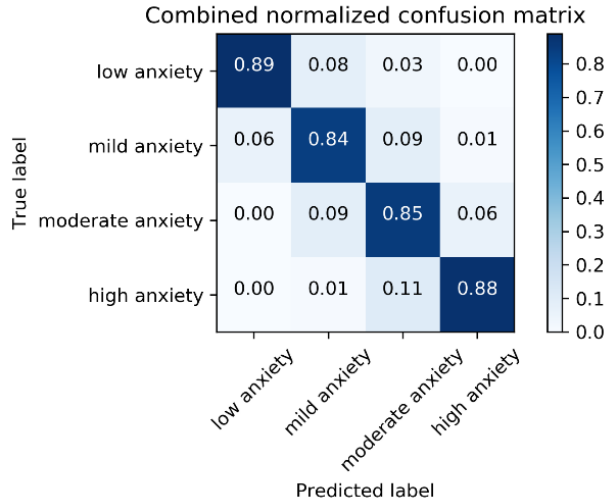


Figure 30. Signal fusion (early) classification normalized confusion matrix

Based on the feature importance analysis, we can see that the early fused model used all the strongest features from every signal. The final anxiety level classification results are detailed in **Table 12**.

Table 12. Anxiety level prediction accuracy by signal type (when using 10x10-fold cross validation)

Signal type	Best accuracy	Windows size
BVP	74.1% (SD = 0.036)	18s
GSR	76.6% (SD = 0.039)	23s
Skin temperature	75.1% (SD = 0.36)	18s
Signal fusion (early)	86.3% (SD = 0.025)	20s
Signal fusion (late)	83.2% (SD = 0.032)	20s

5.6.7 Leave one out subject validation

The standard k-fold validation cannot demonstrate a good validation measurement for a subject to the subject basis as it mixes data from all the participants into separate folds. Therefore, for a subject's independent validation, it is useful to perform 'one-subject-leave-out' validation.

This means that, in our case, we use signals of 29 participants and then test them against one. Our signal fusion (early) model achieved 80.1% accuracy on average with this type of validation. From **Figure 31**, we can see that anxiety prediction accuracy remains high for most of the participants, however, it sharply drops for subjects 11, 15 and 20. This highlights the challenges of predicting the anxiety for each and every subject individually as the results can be impacted by the

subjective nature of the SUDS scale and the challenges of the generalization of the autonomic nervous system response through psychophysiological signals.

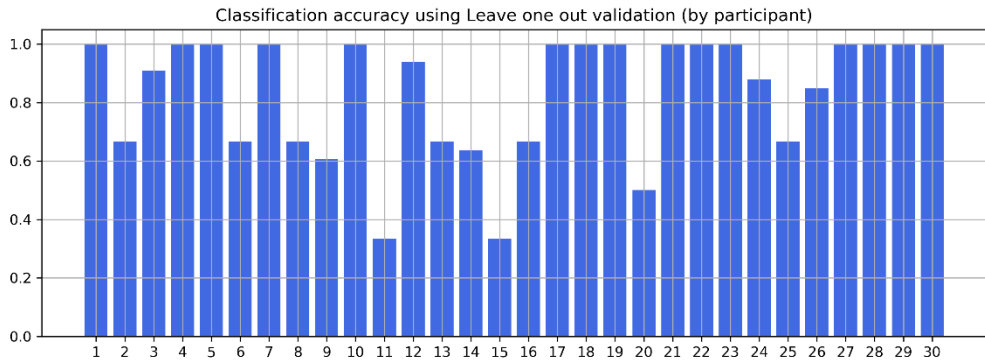


Figure 31. Results of one-subject-out validation

5.6.8 Comparison of results with other works

Multiple studies have been done in the area of stress recognition. The majority of them use various physiological signals in order to detect and classify stress or anxiety, like those that were used in our research. Although, in the course of these studies, anxiety is induced by using various methods, such as completing mental tasks in the laboratory environment, driving, or doing public speaking. Moreover, authors describe anxiety in multiple ways: binary (stressed or relaxed), 3 classes (no anxiety, low anxiety, and high anxiety) or even four classes (no, low, mild, or high anxiety).

We have compared our results to previous research and state that our results are on the same scale as those of other researchers. However, each study and experiment are different, and it is hard to make a strict one-to-one comparison with our research which measures the VRET-induced anxiety.

Thus, we have decided to group the comparison with the mental stress detection system into separate buckets. Firstly, we compare our results with anxiety recognition systems that were trained by using the data collected during the stress periods that were induced in the laboratory; then we discuss the comparison with the studies that used driving-induced stress; and, lastly, we found several studies that measured stress during a public speaking exercise and trained anxiety prediction models for them.

A lot of research studies are using psychophysiological signals collected during the stress induced by solving mental tasks in the laboratory. These tasks usually involve solving logical puzzles, mathematical equations or dealing with time-sensitive challenges. Usually, they have bigger participant groups and are measuring two anxiety classes: one during the relax period (no stress), and one during the solving of the task (stress). While our stress stimuli are slightly different, we still feel that it is important to compare our results with this group as they

represent the majority share of the researched mental stress detection systems (**Table 13**).

Table 13. Comparison with laboratory induced stress recognition systems

Study	Signals	Anxiety stimuli	Method	Subjects	Classes	Accuracy
Vries <i>et al.</i> [85]	GSR, ECG, RV	Tasks in laboratory	LVQ	61	2	88% (10x10-fold)
Xu <i>et al.</i> [127]	GSR, EMG, HR, EEG	Tasks in laboratory	GRNN	39	2	85.2% (leave-1-out)
Akaman dor <i>et al.</i> [128]	GSR, ECG	Tasks in laboratory	SVM	32	2	95.8% (train/test split, subject independent)
Vanitha <i>et al.</i> [129]	EEG	Tasks in laboratory	SVM	6	4	89% (10-fold)
Sandule scu <i>et al.</i> [130]	GSR, BVP	Tasks in laboratory	SVM	5	2	80% (75%/25% split, subject dependent)
Castald o <i>et al.</i> [131]	GSR	Tasks in laboratory	Linear discriminant analysis (LDA)	42	2	98.8% (10-fold subject dependent)
Our system	GSR, BVP, skin temp.	VRET (public speaking)	SVM	30	4	80.1% (leave-1-out) 86.3% (10x10-fold)

Another group of comparison contains research papers which are trying to classify anxiety induced by driving around the city. These researchers are splitting the anxiety level not in 2, but in 3 classes: calm, low stress (low traffic), and high stress (high traffic) situations. It is interesting to compare our results with the datasets which identify the stress level and not only provide binary classification (no stress vs. stress) (**Table 14**).

Table 14. Comparison with driving induced stress recognition systems

Study	Signals	Anxiety stimuli	Method	Subjects	Classes	Accuracy
Chen <i>et al.</i> [132]	GSR, ECG, RV	Driving	Extreme learning machine (ELM)	14	3	99% (cross-drive validation)
Ghaderi <i>et al.</i> [133]	GSR, EMG, ECG	Driving	SVM	7	3	98% (cross validation, no details)
Our system	GSR, BVP, skin temp.	VRET (public speaking)	SVM	30	4	80.1% (leave-1-out) 86.3% (10x10-fold)

Finally, we have found several papers that are presenting anxiety prediction frameworks based on the data collected during public speaking exercises. While they are not based on public speaking in VRET environments, they are the closest to our research (**Table 15**).

Table 15. Comparison with public speaking induced stress recognition systems

Study	Signals	Anxiety stimuli	Method	Subjects	Classes	Accuracy
Zhang <i>et al.</i> [134]	GSR	Public speaking	Neural network (BP)	22	2	86.7% (leave-1-out)
Kurniawian <i>et al.</i> [135]	GSR, speech	Public speaking	SVM	Not specified	3	92% (10x10fold)
Our system	GSR, BVP, skin temp.	VRET (public speaking)	SVM	30	4	80.1% (leave-1-out) 86.3% (10x10-fold)

5.7 Future work

Our further work shall focus on integrating the anxiety level prediction component into the user interface of our VRET system so that the psychologist could easily track and react to the changes of the patient's anxiety during VRET.

Moreover, our next development steps of the VRET system shall focus on extracting and delivering biofeedback signal features in the real time on the therapist's dashboard. This will help psychologists to evaluate the patient's psychological state during the VRET session and increase or decrease the intensity

of a simulated virtual environment. Also, anxiety level prediction during public anxiety treatment sessions could benefit from temporal processing as an additional source of auditory signals. While in our experiment we were not recording the speech itself, we strongly believe that future research should consider recording participants' speech patterns and add them as an additional modality for the feature extraction phase.

Additionally, a self-treatment online module could be developed so that patients will be able to use the system without interaction with the therapist. Finally, the system will be prepared to be used for a fully randomized clinical trial which will include the control group.

5.8 Conclusions of the Chapter

In this chapter, we have described the experiment which involved 30 participants going through public speaking intervention while using our VRET system. During the experiment, we collected the subjects' psychophysiological signals and used them for training anxiety recognition models. We have investigated the signal's windows size, the features of psychophysiological signals, and the signal fusion that provides the best anxiety recognition accuracy for a VRET system.

The combination of anxiety detection and VRET demonstrates a possibility to measure mental anxiety by adding real-time stress recognition capabilities. Our best classification model was capable to predict a four-class anxiety level with high 80.1% accuracy (when employing one-subject-out validation) and 86.3% (when employing 10x10-fold cross-validation). This should encourage other researchers and VRET system developers to invest time into this kind of addition of psychophysiological signal analysis for their VR solutions. However, in our work, we have also investigated various limitations of VRET solutions and mental stress recognition systems that must be considered for further research. This is also highlighted through one-subject-out validation as the anxiety prediction of some subjects yielded lower results.

We have also investigated the impact of the signal window size on the anxiety prediction accuracy as different stimuli can have different emotional activation lengths. Our results showed that the best performing VRET stimuli-based models featured between 20s and 25s window length. We have reconfirmed the results by many previous studies that fused signal datasets are performing better than stand-alone signals and showed that this applies for VRET based anxiety stimuli as well. The classification model with signal early fusion performed by 9.7% better than the GSR-based model, by 11.2% better than the skin temperature-based model, and by 12.2% better than the BVP-based model.

While there are a lot of studies covering VRET systems and mental anxiety recognition solutions, our designed system combines these fields together. Our goal is to demonstrate that there exists a potential way of estimating the anxiety level of the patient who is undergoing VRET sessions. This information could help to

improve VRET systems which could react and change VR session scenarios in real-time (under the supervision of a therapist) based on the patient's mental stress.

The limitations of our anxiety recognition methods include the reliance on the subjective anxiety scale (SUDS) which is used to add classification labels to our dataset. Additionally, the current experiment was limited to a single treatment session per participant, while, usually, anxiety treatment takes multiple sessions.

Lastly, it can be concluded that psychophysiological signals collected through a wearable device can be used to train a model for anxiety recognition during VRET sessions. Moreover, it is feasible to determine gradual changes in patient anxiety, and this may help the psychologist balance and adapt the VRET session itself.

6 CONCLUSIONS

1. The proposed VRET system based on the VRaaS model allows the therapist to interact with the 3D environment and treatment scenarios in the real-time and to change and adapt virtual environments during VRET sessions for personalized treatment.
2. Our VRET system with the four-class anxiety (low, mild, moderate, and high) recognition component provides the possibility to the therapist to react to the patient's anxiety changes in the real-time thus enabling the therapist to lower or to increase the intensity of the virtual reality therapy scenario based on the observed patient's anxiety level.
3. The proposed four-class anxiety classification method based on the features extracted from psychophysiological signals can recognize the patient's anxiety during a VRET session with the accuracy of 86.3% when using 10x10-fold cross-validation and with 80.1% accuracy when using subject-independent leave one out cross-validation.

REFERENCES

- [1] P. Lindner, A. Miloff, E. Zetterlund, L. Reuterskiöld, G. Andersson, and P. Carlbring, "Attitudes Toward and Familiarity With Virtual Reality Therapy Among Practicing Cognitive Behavior Therapists: A Cross-Sectional Survey Study in the Era of Consumer VR Platforms," *Front. Psychol.*, vol. 10, p. 176, Feb. 2019, doi: 10.3389/fpsyg.2019.00176.
- [2] E. Carl *et al.*, "Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials," *J. Anxiety Disord.*, vol. 61, pp. 27–36, Jan. 2019, doi: 10.1016/J.JANXDIS.2018.08.003.
- [3] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, "Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication," *Arch. Gen. Psychiatry*, vol. 62, no. 6, p. 593, Jun. 2005, doi: 10.1001/archpsyc.62.6.593.
- [4] B. Bandelow and S. Michaelis, "Epidemiology of anxiety disorders in the 21st century.," *Dialogues Clin. Neurosci.*, vol. 17, no. 3, pp. 327–35, Sep. 2015, Accessed: Jun. 06, 2019. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26487813>.
- [5] V. Raudonis, R. Maskeliūnas, K. Stankevičius, and R. Damaševičius, "Gender, Age, Colour, Position and Stress: How They Influence Attention at Workplace?," Springer, Cham, 2017, pp. 248–264.
- [6] J. K. Carpenter, M. Pinaire, and S. G. Hofmann, "From Extinction Learning to Anxiety Treatment: Mind the Gap.," *Brain Sci.*, vol. 9, no. 7, Jul. 2019, doi: 10.3390/brainsci9070164.
- [7] F. Scibelli, A. Troncone, L. Likforman-Sulem, A. Vinciarelli, and A. Esposito, "How Major Depressive Disorder Affects the Ability to Decode Multimodal Dynamic Emotional Stimuli," *Front. ICT*, vol. 3, p. 16, Sep. 2016, doi: 10.3389/fict.2016.00016.
- [8] J. K. Carpenter, L. A. Andrews, S. M. Witcraft, M. B. Powers, J. A. J. Smits, and S. G. Hofmann, "Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials," *Depress. Anxiety*, vol. 35, no. 6, pp. 502–514, Jun. 2018, doi: 10.1002/da.22728.
- [9] R. A. McCann *et al.*, "Virtual reality exposure therapy for the treatment of anxiety disorders: An evaluation of research quality," *J. Anxiety Disord.*, vol. 28, no. 6, pp. 625–631, Aug. 2014, doi: 10.1016/J.JANXDIS.2014.05.010.
- [10] H. K. Hood and M. M. Antony, "Evidence-Based Assessment and Treatment of Specific Phobias in Adults," Springer, New York, NY, 2012, pp. 19–42.
- [11] Q. A. Le, J. N. Doctor, L. A. Zoellner, and N. C. Feeny, "Cost-Effectiveness of Prolonged Exposure Therapy Versus Pharmacotherapy and Treatment Choice in Posttraumatic Stress Disorder (the Optimizing PTSD Treatment Trial)," *J. Clin. Psychiatry*, vol. 75, no. 03, pp. 222–230, Mar. 2014, doi: 10.4088/JCP.13m08719.

- [12] D. C. Beidel *et al.*, “Trauma management therapy with virtual-reality augmented exposure therapy for combat-related PTSD: A randomized controlled trial,” *J. Anxiety Disord.*, vol. 61, pp. 64–74, Jan. 2019, doi: 10.1016/J.JANXDIS.2017.08.005.
- [13] R. Buzys, R. Maskeliūnas, R. Damaševičius, T. Sidekerskienė, M. Woźniak, and W. Wei, “Cloudification of Virtual Reality Gliding Simulation Game,” *Information*, vol. 9, no. 12, p. 293, Nov. 2018, doi: 10.3390/info9120293.
- [14] A. A. Benbow and P. L. Anderson, “A meta-analytic examination of attrition in virtual reality exposure therapy for anxiety disorders,” *J. Anxiety Disord.*, vol. 61, pp. 18–26, Jan. 2019, doi: 10.1016/J.JANXDIS.2018.06.006.
- [15] G. Mertens, P. Wagenveld, and I. M. Engelhard, “Cue conditioning using a virtual spider discriminates between high and low spider fearful individuals,” *Comput. Human Behav.*, vol. 91, pp. 192–200, Feb. 2019, doi: 10.1016/J.CHB.2018.10.006.
- [16] S. D. Norrholm *et al.*, “Baseline psychophysiological and cortisol reactivity as a predictor of PTSD treatment outcome in virtual reality exposure therapy,” *Behav. Res. Ther.*, vol. 82, pp. 28–37, Jul. 2016, doi: 10.1016/J.BRAT.2016.05.002.
- [17] R. Maskeliūnas, T. Blažauskas, and R. Damaševičius, “Depression Behavior Detection Model Based on Participation in Serious Games,” Springer, Cham, 2017, pp. 423–434.
- [18] R. W. Picard, *Affective computing*. MIT press, 2000.
- [19] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: analysis of affective physiological state,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001, doi: 10.1109/34.954607.
- [20] E. Vaškevičius, A. Vidugirienė, and V. Kaminskas, “Identification of Human Response to Virtual 3D Face Stimuli,” *Inf. Technol. Control*, vol. 43, no. 1, Mar. 2014, doi: 10.5755/j01.itc.43.1.5927.
- [21] S. Katsigiannis and N. Ramzan, “DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices,” *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 1, pp. 98–107, Jan. 2018, doi: 10.1109/JBHI.2017.2688239.
- [22] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, “DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses,” *IEEE Trans. Affect. Comput.*, vol. 6, no. 3, pp. 209–222, Jul. 2015, doi: 10.1109/TAFFC.2015.2392932.
- [23] J. Moraes, M. Rocha, G. Vasconcelos, J. Vasconcelos Filho, V. de Albuquerque, and A. Alexandria, “Advances in Photoplethysmography Signal Analysis for Biomedical Applications,” *Sensors*, vol. 18, no. 6, p. 1894, Jun. 2018, doi: 10.3390/s18061894.
- [24] N. Murali Krishna *et al.*, “An Efficient Mixture Model Approach in Brain-Machine Interface Systems for Extracting the Psychological Status of Mentally Impaired Persons Using EEG Signals,” *IEEE Access*, vol. 7, pp. 77905–77914, 2019, doi: 10.1109/ACCESS.2019.2922047.

- [25] M. E. Owens and D. C. Beidel, "Can Virtual Reality Effectively Elicit Distress Associated with Social Anxiety Disorder?," *J. Psychopathol. Behav. Assess.*, vol. 37, no. 2, pp. 296–305, Jun. 2015, doi: 10.1007/s10862-014-9454-x.
- [26] L. R. Valmaggia, L. Latif, M. J. Kempton, and M. Rus-Calafell, "Virtual reality in the psychological treatment for mental health problems: A systematic review of recent evidence.," *Psychiatry Res.*, vol. 236, pp. 189–95, Feb. 2016, doi: 10.1016/j.psychres.2016.01.015.
- [27] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shibana, and A. Mählberger, "The impact of perception and presence on emotional reactions: a review of research in virtual reality," *Front. Psychol.*, vol. 6, p. 26, Jan. 2015, doi: 10.3389/fpsyg.2015.00026.
- [28] K. R. Gujjar, R. Sharma, and A. De Jongh, "Virtual reality exposure therapy for treatment of dental phobia," *Dent. Update*, vol. 44, no. 5, pp. 423–435, May 2017, doi: 10.12968/denu.2017.44.5.423.
- [29] A.-D. Tudor, S. Poeschl, and N. Doering, "Virtual audience customization for public speaking training procedures," in *2013 IEEE Virtual Reality (VR)*, Mar. 2013, pp. 61–62, doi: 10.1109/VR.2013.6549363.
- [30] C. Botella, J. Fernández-Álvarez, V. Guillén, A. García-Palacios, and R. Baños, "Recent Progress in Virtual Reality Exposure Therapy for Phobias: A Systematic Review," *Curr. Psychiatry Rep.*, vol. 19, no. 7, p. 42, Jul. 2017, doi: 10.1007/s11920-017-0788-4.
- [31] S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 444–469, Apr. 2019, doi: 10.1016/J.BBE.2019.01.004.
- [32] S. R. Harris, R. L. Kemmerling, and M. M. North, "Brief Virtual Reality Therapy for Public Speaking Anxiety," *CyberPsychology Behav.*, vol. 5, no. 6, pp. 543–550, Dec. 2002, doi: 10.1089/109493102321018187.
- [33] M. C. Mishkind, A. M. Norr, A. C. Katz, and G. M. Reger, "Review of Virtual Reality Treatment in Psychiatry: Evidence Versus Current Diffusion and Use," *Curr. Psychiatry Rep.*, vol. 19, no. 11, p. 80, Nov. 2017, doi: 10.1007/s11920-017-0836-0.
- [34] R. A. I. Cardoso, O. A. David, and D. O. David, "Virtual reality exposure therapy in flight anxiety: A quantitative meta-analysis," *Comput. Human Behav.*, vol. 72, pp. 371–380, Jul. 2017, doi: 10.1016/J.CHB.2017.03.007.
- [35] S. Scapin, M. E. Echevarría-Guanilo, P. R. Boeira Fuculo Junior, N. Gonçalves, P. K. Rocha, and R. Coimbra, "Virtual Reality in the treatment of burn patients: A systematic review," *Burns*, vol. 44, no. 6, pp. 1403–1416, Sep. 2018, doi: 10.1016/J.BURNS.2017.11.002.
- [36] C. Botella, J. Fernández-Álvarez, V. Guillén, A. García-Palacios, and R. Baños, "Recent Progress in Virtual Reality Exposure Therapy for Phobias: A Systematic Review," *Curr. Psychiatry Rep.*, vol. 19, no. 7, p. 42, Jul. 2017, doi: 10.1007/s11920-017-0788-4.
- [37] E. Carl *et al.*, "Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials," *J. Anxiety*

- Disord.*, Aug. 2018, doi: 10.1016/J.JANXDIS.2018.08.003.
- [38] F. Levy, P. Leboucher, G. Rautureau, and R. Jouvent, “E-virtual reality exposure therapy in acrophobia: A pilot study,” *J. Telemed. Telecare*, vol. 22, no. 4, pp. 215–220, Jun. 2016, doi: 10.1177/1357633X15598243.
 - [39] E. M. Suyanto, D. Angkasa, H. Turaga, and R. Sutoyo, “Overcome Acrophobia with the Help of Virtual Reality and Kinect Technology,” *Procedia Comput. Sci.*, vol. 116, pp. 476–483, Jan. 2017, doi: 10.1016/J.PROCS.2017.10.062.
 - [40] S. Shunnaq and M. Raeder, “VirtualPhobia: A Model for Virtual Therapy of Phobias,” in *2016 XVIII Symposium on Virtual and Augmented Reality (SVR)*, Jun. 2016, pp. 59–63, doi: 10.1109/SVR.2016.20.
 - [41] S. A. M. Rauch *et al.*, “In session extinction and outcome in Virtual Reality Exposure Therapy for PTSD,” *Behav. Res. Ther.*, vol. 109, pp. 1–9, Oct. 2018, doi: 10.1016/J.BRAT.2018.07.003.
 - [42] H. M. Zinzow *et al.*, “Virtual Reality and Cognitive-Behavioral Therapy for Driving Anxiety and Aggression in Veterans: A Pilot Study,” *Cogn. Behav. Pract.*, vol. 25, no. 2, pp. 296–309, May 2018, doi: 10.1016/J.CBPRA.2017.09.002.
 - [43] A. Marquardt, C. Trepkowski, J. Maiero, E. Kruijff, and A. Hinkenian, “Multisensory Virtual Reality Exposure Therapy,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Mar. 2018, pp. 769–770, doi: 10.1109/VR.2018.8446553.
 - [44] M. Christofi and D. Michael-Grigoriou, “Virtual environments design assessment for the treatment of claustrophobia,” in *2016 22nd International Conference on Virtual System & Multimedia (VSMM)*, Oct. 2016, pp. 1–8, doi: 10.1109/VSMM.2016.7863215.
 - [45] D. C. Beidel *et al.*, “Trauma management therapy with virtual-reality augmented exposure therapy for combat-related PTSD: A randomized controlled trial,” *J. Anxiety Disord.*, vol. 61, pp. 64–74, Jan. 2019, doi: 10.1016/J.JANXDIS.2017.08.005.
 - [46] D. Freeman *et al.*, “Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial,” *The Lancet Psychiatry*, vol. 5, no. 8, pp. 625–632, Aug. 2018, doi: 10.1016/S2215-0366(18)30226-8.
 - [47] T. Manju, S. Padmavathi, and D. Tamilselvi, “A Rehabilitation Therapy for Autism Spectrum Disorder Using Virtual Reality,” 2018, pp. 328–336.
 - [48] J. Diemer, N. Lohkamp, A. Mühlberger, and P. Zwanzger, “Fear and physiological arousal during a virtual height challenge—effects in patients with acrophobia and healthy controls,” *J. Anxiety Disord.*, vol. 37, pp. 30–39, Jan. 2016, doi: 10.1016/j.janxdis.2015.10.007.
 - [49] K. Raghav, A. Van Wijk, F. Abdullah, M. N. Islam, M. Bernatchez, and A. De Jongh, “Efficacy of virtual reality exposure therapy for treatment of dental phobia: a randomized control trial,” *BMC Oral Health*, vol. 16, no. 1, p. 25, Dec. 2016, doi: 10.1186/s12903-016-0186-z.
 - [50] A. Miloff, P. Lindner, W. Hamilton, L. Reuterskiöld, G. Andersson, and P.

- Carlbring, "Single-session gamified virtual reality exposure therapy for spider phobia vs. traditional exposure therapy: study protocol for a randomized controlled non-inferiority trial," *Trials*, vol. 17, no. 1, p. 60, Dec. 2016, doi: 10.1186/s13063-016-1171-1.
- [51] Y. Shibani, J. Diemer, J. Müller, J. Brütting-Schick, P. Pauli, and A. Mühlberger, "Diaphragmatic breathing during virtual reality exposure therapy for aviophobia: functional coping strategy or avoidance behavior? a pilot study," *BMC Psychiatry*, vol. 17, no. 1, p. 29, Dec. 2017, doi: 10.1186/s12888-016-1181-2.
- [52] S. Skulimowski and M. Badurowicz, "Wearable sensors as feedback method in virtual reality anti-stress therapy," in *2017 International Conference on Electromagnetic Devices and Processes in Environment Protection with Seminar Applications of Superconductors (ELMECO & AoS)*, Dec. 2017, pp. 1–4, doi: 10.1109/ELMECO.2017.8267716.
- [53] E. Seol *et al.*, "Drop the beat": Virtual reality based mindfulness and cognitive behavioral therapy for panic disorder — A pilot study," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17*, 2017, pp. 1–3, doi: 10.1145/3139131.3141199.
- [54] B. K. Wiederhold, I. T. Miller, and M. D. Wiederhold, "Using Virtual Reality to Mobilize Health Care: Mobile Virtual Reality Technology for Attenuation of Anxiety and Pain," *IEEE Consum. Electron. Mag.*, vol. 7, no. 1, pp. 106–109, Jan. 2018, doi: 10.1109/MCE.2017.2715365.
- [55] A. A. Benbow and P. L. Anderson, "A meta-analytic examination of attrition in virtual reality exposure therapy for anxiety disorders," *J. Anxiety Disord.*, vol. 61, pp. 18–26, Jan. 2019, doi: 10.1016/J.JANXDIS.2018.06.006.
- [56] P. Bun, F. Gorski, D. Grajewski, R. Wichniarek, and P. Zawadzki, "Low – Cost Devices Used in Virtual Reality Exposure Therapy," *Procedia Comput. Sci.*, vol. 104, pp. 445–451, Jan. 2017, doi: 10.1016/J.PROCS.2017.01.158.
- [57] E. Bekele, D. Bian, J. Peterman, S. Park, and N. Sarkar, "Design of a Virtual Reality System for Affect Analysis in Facial Expressions (VR-SAAFE); Application to Schizophrenia," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 739–749, Jun. 2017, doi: 10.1109/TNSRE.2016.2591556.
- [58] A. Dhiman, D. Solanki, A. Bhasin, A. Das, and U. Lahiri, "An intelligent, adaptive, performance-sensitive, and virtual reality-based gaming platform for the upper limb," *Comput. Animat. Virtual Worlds*, vol. 29, no. 2, p. e1800, Mar. 2018, doi: 10.1002/cav.1800.
- [59] W.-P. Brinkman and N. Ter Heijden, "Design and Evaluation of a Virtual Reality Exposure Therapy System with Automatic Free Speech Interaction," 2011. Accessed: Sep. 01, 2019. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.208.3293&rep=rep1&type=pdf>.
- [60] M. Ben-Moussa, M. Rubo, C. Debracque, and W.-G. Lange, "DJINNI: A Novel Technology Supported Exposure Therapy Paradigm for SAD Combining Virtual Reality and Augmented Reality," *Front. Psychiatry*, vol. 8, Apr. 2017, doi: 10.3389/fpsy.2017.00026.

- [61] J. Marín-Morales *et al.*, “Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors,” *Sci. Rep.*, vol. 8, no. 1, p. 13657, Dec. 2018, doi: 10.1038/s41598-018-32063-4.
- [62] I. Mavridou, E. Seiss, T. Kostoulas, C. Nduka, and E. Balaguer-Ballester, “Towards an effective arousal detection system for virtual reality,” in *Proceedings of the Workshop on Human-Habitat for Health (H3) Human-Habitat Multimodal Interaction for Promoting Health and Well-Being in the Internet of Things Era - H3 '18*, 2018, pp. 1–6, doi: 10.1145/3279963.3279969.
- [63] M. Moghimi, R. Stone, and P. Rotshtein, “Affective Recognition in Dynamic and Interactive Virtual Environments,” *IEEE Trans. Affect. Comput.*, pp. 1–1, 2017, doi: 10.1109/TAFFC.2017.2764896.
- [64] M. Moghim, R. Stone, P. Rotshtein, and N. Cooke, “Adaptive virtual environments: A physiological feedback HCI system concept,” in *2015 7th Computer Science and Electronic Engineering Conference (CEEC)*, Sep. 2015, pp. 123–128, doi: 10.1109/CEEC.2015.7332711.
- [65] C. Christou, K. Herakleous, A. Tzanavari, and C. Poullis, “Psychophysiological responses to virtual crowds: Implications for wearable computing,” in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Sep. 2015, pp. 35–41, doi: 10.1109/ACII.2015.7344548.
- [66] D. Wu *et al.*, “Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task,” *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 109–118, Jul. 2010, doi: 10.1109/T-AFFC.2010.12.
- [67] M. A. Zafar, B. Ahmed, R. Al-Rihawi, and R. Gutierrez-Osuna, “Gaming away stress: Using biofeedback games to learn paced breathing,” *IEEE Trans. Affect. Comput.*, pp. 1–1, 2018, doi: 10.1109/TAFFC.2018.2816945.
- [68] M. L. R. Menezes *et al.*, “Towards emotion recognition for virtual environments: an evaluation of eeg features on benchmark dataset,” *Pers. Ubiquitous Comput.*, vol. 21, no. 6, pp. 1003–1013, Dec. 2017, doi: 10.1007/s00779-017-1072-7.
- [69] S. Koelstra *et al.*, “DEAP: A Database for Emotion Analysis ;Using Physiological Signals,” *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012, doi: 10.1109/T-AFFC.2011.15.
- [70] Y. S. Can, B. Arnrich, and C. Ersoy, “Stress detection in daily life scenarios using smart phones and wearable sensors: A survey,” *J. Biomed. Inform.*, vol. 92, p. 103139, Apr. 2019, doi: 10.1016/J.JBI.2019.103139.
- [71] G. Walkom, “Virtual Reality Exposure Therapy: To Benefit Those Who Stutter and Treat Social Anxiety,” in *2016 International Conference on Interactive Technologies and Games (ITAG)*, Oct. 2016, pp. 36–41, doi: 10.1109/iTAG.2016.13.
- [72] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, “Stress detection from speech and Galvanic Skin Response signals,” in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, Jun.

- 2013, pp. 209–214, doi: 10.1109/CBMS.2013.6627790.
- [73] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, “Continuous stress detection using a wrist device,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, 2016, pp. 1185–1193, doi: 10.1145/2968219.2968306.
 - [74] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, “Monitoring stress with a wrist device using context,” *J. Biomed. Inform.*, vol. 73, pp. 159–170, Sep. 2017, doi: 10.1016/J.JBI.2017.08.006.
 - [75] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, “The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain,” *J. Psychiatry Neurosci.*, vol. 30, no. 5, pp. 319–25, Sep. 2005, Accessed: Jun. 07, 2019. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16151536>.
 - [76] D. Das, T. Bhattacharjee, S. Datta, A. D. Choudhury, P. Das, and A. Pal, “Classification and quantitative estimation of cognitive stress from in-game keystroke analysis using EEG and GSR,” in *2017 IEEE Life Sciences Conference (LSC)*, Dec. 2017, pp. 286–291, doi: 10.1109/LSC.2017.8268199.
 - [77] R. T. Azevedo, N. Bennett, A. Bilicki, J. Hooper, F. Markopoulou, and M. Tsakiris, “The calming effect of a new wearable device during the anticipation of public speech,” *Sci. Rep.*, vol. 7, no. 1, p. 2285, Dec. 2017, doi: 10.1038/s41598-017-02274-2.
 - [78] R. Kuboi, K. Jimi, M. Inoue, and I. Matsunami, “Estimating stress states using respiratory monitoring by 24GHz wideband radar,” in *2016 IEEE 5th Global Conference on Consumer Electronics*, Oct. 2016, pp. 1–4, doi: 10.1109/GCCE.2016.7800499.
 - [79] N. Sowmya, S. Srivarshini, N. Shanmathi, and R. Menaka, “Stress Diagonisis Using EMG Signals,” in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Mar. 2018, pp. 1–4, doi: 10.1109/ICCTCT.2018.8551028.
 - [80] J. Hayano, T. Tanabiki, S. Iwata, K. Abe, and E. Yuda, “Estimation of Emotions by Wearable Biometric Sensors Under Daily Activities,” in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, Oct. 2018, pp. 240–241, doi: 10.1109/GCCE.2018.8574758.
 - [81] A. Hernando *et al.*, “Inclusion of Respiratory Frequency Information in Heart Rate Variability Analysis for Stress Assessment,” *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 4, pp. 1016–1025, Jul. 2016, doi: 10.1109/JBHI.2016.2553578.
 - [82] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, “Validation of the Empatica E4 wristband,” in *2016 IEEE EMBS International Student Conference (ISC)*, May 2016, pp. 1–4, doi: 10.1109/EMBSISC.2016.7508621.
 - [83] H. Sharma, “Heart rate extraction from PPG signals using variational mode

- decomposition,” *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 75–86, Jan. 2019, doi: 10.1016/j.bbe.2018.11.001.
- [84] F. Delmastro, F. Di Martino, and C. Dolciotti, “Physiological Impact of Vibro-Acoustic Therapy on Stress and Emotions through Wearable Sensors,” in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Mar. 2018, pp. 621–626, doi: 10.1109/PERCOMW.2018.8480170.
- [85] J. J. G. (Gert-J. de Vries, S. C. Pauws, and M. Biehl, “Insightful stress detection from physiology modalities using Learning Vector Quantization,” *Neurocomputing*, vol. 151, pp. 873–882, Mar. 2015, doi: 10.1016/J.NEUCOM.2014.10.008.
- [86] T. A. Furukawa *et al.*, “Evidence-based guidelines for interpretation of the Panic Disorder Severity Scale,” *Depress. Anxiety*, vol. 26, no. 10, pp. 922–9, 2009, doi: 10.1002/da.20532.
- [87] D. J. A. Dozois and H. A. Westra, “Development of the Anxiety Change Expectancy Scale (ACES) and validation in college, community, and clinical samples,” *Behav. Res. Ther.*, vol. 43, no. 12, pp. 1655–1672, Dec. 2005, doi: 10.1016/j.brat.2004.12.001.
- [88] L. A. Rutter and T. A. Brown, “Psychometric Properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in Outpatients with Anxiety and Mood Disorders,” *J. Psychopathol. Behav. Assess.*, vol. 39, no. 1, pp. 140–146, Mar. 2017, doi: 10.1007/s10862-016-9571-9.
- [89] “An Inventory for Measuring Clinical Anxiety: Psychometric Properties,” 1988. Accessed: Aug. 27, 2019. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.4319&rep=rep1&type=pdf>.
- [90] J. Wolpe, *The practice of behavior therapy*. New York, New York, USA: Pergamon Press, 1969.
- [91] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, “ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors,” *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 147–160, Apr. 2018, doi: 10.1109/TAFFC.2016.2625250.
- [92] M. Liu, D. Fan, X. Zhang, and X. Gong, “Human Emotion Recognition Based on Galvanic Skin Response Signal Feature Selection and SVM,” in *2016 International Conference on Smart City and Systems Engineering (ICSCSE)*, Nov. 2016, pp. 157–160, doi: 10.1109/ICSCSE.2016.0051.
- [93] Q. Zhang, X. Lai, and G. Liu, “Emotion Recognition of GSR Based on an Improved Quantum Neural Network,” in *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Aug. 2016, pp. 488–492, doi: 10.1109/IHMSC.2016.66.
- [94] R. W. Picard, “Affective Computing,” Accessed: Jun. 06, 2019. [Online]. Available: <http://www.media.mit.edu/~picard/>.
- [95] N. P. Jagini and R. R. Rao, “Exploring emotion specific features for emotion recognition system using PCA approach,” in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Jun. 2017, pp. 58–

- 62, doi: 10.1109/ICCONS.2017.8250527.
- [96] I. Daly *et al.*, “Identifying music-induced emotions from EEG for use in brain-computer music interfacing,” in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Sep. 2015, pp. 923–929, doi: 10.1109/ACII.2015.7344685.
 - [97] S. D. Norrholm *et al.*, “Baseline psychophysiological and cortisol reactivity as a predictor of PTSD treatment outcome in virtual reality exposure therapy,” *Behav. Res. Ther.*, vol. 82, pp. 28–37, Jul. 2016, doi: 10.1016/J.BRAT.2016.05.002.
 - [98] M. Fusaro, G. Tieri, and S. M. Aglioti, “Seeing pain and pleasure on self and others: behavioral and psychophysiological reactivity in immersive virtual reality,” *J. Neurophysiol.*, vol. 116, no. 6, pp. 2656–2662, Dec. 2016, doi: 10.1152/jn.00489.2016.
 - [99] L. Detez *et al.*, “A Psychophysiological and Behavioural Study of Slot Machine Near-Misses Using Immersive Virtual Reality,” *J. Gambl. Stud.*, vol. 35, no. 3, pp. 929–944, Sep. 2019, doi: 10.1007/s10899-018-09822-z.
 - [100] S. Notzon *et al.*, “Psychophysiological effects of an iTBS modulated virtual reality challenge including participants with spider phobia,” *Biol. Psychol.*, vol. 112, pp. 66–76, Dec. 2015, doi: 10.1016/J.BIOPSYCHO.2015.10.003.
 - [101] C. Breuninger, D. M. Sláma, M. Krämer, J. Schmitz, and B. Tuschen-Caffier, “Psychophysiological Reactivity, Interoception and Emotion Regulation in Patients with Agoraphobia During Virtual Reality Anxiety Induction,” *Cognit. Ther. Res.*, vol. 41, no. 2, pp. 193–205, Apr. 2017, doi: 10.1007/s10608-016-9814-9.
 - [102] S. Kuriakose and U. Lahiri, “Understanding the Psycho-Physiological Implications of Interaction With a Virtual Reality-Based System in Adolescents With Autism: A Feasibility Study,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 665–675, Jul. 2015, doi: 10.1109/TNSRE.2015.2393891.
 - [103] T. Gao, T. Zhang, L. Zhu, Y. Gao, and L. Qiu, “Exploring Psychophysiological Restoration and Individual Preference in the Different Environments Based on Virtual Reality,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 17, p. 3102, Aug. 2019, doi: 10.3390/ijerph16173102.
 - [104] G. Czarnek, P. Strojny, A. Strojny, and M. Richter, “Assessing Engagement during Rescue Operation Simulated in Virtual Reality: A Psychophysiological Study,” *Int. J. Human–Computer Interact.*, pp. 1–13, Aug. 2019, doi: 10.1080/10447318.2019.1655905.
 - [105] F. G. Esteves, A. P. Claudio, M. B. Carmo, and M. A. Gaspar, “Psychophysiological responses to a Virtual Reality scenario for the treatment of Social Anxiety,” *Int. J. Psychophysiol.*, vol. 108, p. 138, Oct. 2016, doi: 10.1016/J.IJPSYCHO.2016.07.403.
 - [106] J. Salkevicius and L. Navickas, “Battling the Fear of Public Speaking: Designing Software as a Service Solution for a Virtual Reality Therapy,” in *2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, Aug. 2018, pp. 209–213, doi: 10.1109/W-

- FiCloud.2018.00040.
- [107] J. Šalkevičius, A. Miškinytė, and L. Navickas, "Cloud Based Virtual Reality Exposure Therapy Service for Public Speaking Anxiety," *Information*, vol. 10, no. 2, p. 62, Feb. 2019, doi: 10.3390/info10020062.
 - [108] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy Array: A Structure for Efficient Numerical Computation," *Comput. Sci. Eng.*, vol. 13, no. 2, pp. 22–30, Mar. 2011, doi: 10.1109/MCSE.2011.37.
 - [109] C. Carreiras *et al.*, "BioSPPy: Biosignal Processing in Python." [Online]. Available: <https://github.com/PIA-Group/BioSPPy/>.
 - [110] P. van Gent, H. Farah, N. Nes, and B. van Arem, "Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data," *6th Humanist Conf.*, 2018, Accessed: Jun. 25, 2019. [Online]. Available: <https://repository.tudelft.nl/islandora/object/uuid%3A5c638e14-d249-4116-aa05-2e566cf3df02>.
 - [111] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," Jan. 2012, Accessed: Aug. 25, 2019. [Online]. Available: <http://arxiv.org/abs/1201.0490>.
 - [112] E. M. Bartholomay and D. D. Houlihan, "Public Speaking Anxiety Scale: Preliminary psychometric data and scale validation," *Pers. Individ. Dif.*, vol. 94, pp. 211–215, May 2016, doi: 10.1016/J.PAID.2016.01.026.
 - [113] J. C. McCroskey, M. J. Beatty, P. Kearney, and T. G. Plax, "The content validity of the PRCA-24 as a measure of communication apprehension across communication contexts," *Commun. Q.*, vol. 33, no. 3, pp. 165–173, Jun. 1985, doi: 10.1080/01463378509369595.
 - [114] Y. Cho, J. A. J. Smits, and M. J. Telch, "The Speech Anxiety Thoughts Inventory: scale development and preliminary psychometric data," *Behav. Res. Ther.*, vol. 42, no. 1, pp. 13–25, Jan. 2004, doi: 10.1016/S0005-7967(03)00067-6.
 - [115] D.-P. Pertaub, M. Slater, and C. Barker, "An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience," *Presence Teleoperators Virtual Environ.*, vol. 11, no. 1, pp. 68–78, Feb. 2002, doi: 10.1162/105474602317343668.
 - [116] M. Krijn, P. M. . Emmelkamp, R. Biemond, C. de Wilde de Ligny, M. J. Schuemie, and C. A. P. . van der Mast, "Treatment of acrophobia in virtual reality: The role of immersion and presence," *Behav. Res. Ther.*, vol. 42, no. 2, pp. 229–239, Feb. 2004, doi: 10.1016/S0005-7967(03)00139-6.
 - [117] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion Based Music Recommendation System Using Wearable Physiological Sensors," *IEEE Trans. Consum. Electron.*, vol. 64, no. 2, pp. 196–203, May 2018, doi: 10.1109/TCE.2018.2844736.
 - [118] W. Wen, G. Liu, N. Cheng, J. Wei, P. Shangguan, and W. Huang, "Emotion Recognition Based on Multi-Variant Correlation of Physiological Signals," *IEEE Trans. Affect. Comput.*, vol. 5, no. 2, pp. 126–140, Apr. 2014, doi: 10.1109/TAFFC.2014.2327617.

- [119] M. Garbarino, M. Lai, S. Tognetti, R. Picard, and D. Bender, "Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming healthcare through innovations in mobile and wireless technologies,"* 2014, doi: 10.4108/icst.mobihealth.2014.257418.
- [120] K. R. Gujjar, A. van Wijk, R. Kumar, and A. de Jongh, "Efficacy of virtual reality exposure therapy for the treatment of dental phobia in adults: A randomized controlled trial," *J. Anxiety Disord.*, vol. 62, pp. 100–108, Mar. 2019, doi: 10.1016/j.janxdis.2018.12.001.
- [121] A. Garcia-Palacios, H. Hoffman, A. Carlin, T. . Furness, and C. Botella, "Virtual reality in the treatment of spider phobia: a controlled study," *Behav. Res. Ther.*, vol. 40, no. 9, pp. 983–993, Sep. 2002, doi: 10.1016/S0005-7967(01)00068-7.
- [122] A. Mühlberger, M. J. Herrmann, G. Wiedemann, H. Ellgring, and P. Pauli, "Repeated exposure of flight phobics to flights in virtual reality," *Behav. Res. Ther.*, vol. 39, no. 9, pp. 1033–1050, Sep. 2001, doi: 10.1016/S0005-7967(00)00076-0.
- [123] E. E. Osuna, R. Freund, and F. Girosi, "Support Vector Machines: Training and Applications," Mar. 1997.
- [124] W.-H. Cao, J.-P. Xu, and Z.-T. Liu, "Speaker-independent speech emotion recognition based on random forest feature selection algorithm," in *2017 36th Chinese Control Conference (CCC)*, Jul. 2017, pp. 10995–10998, doi: 10.23919/ChiCC.2017.8029112.
- [125] R. Sioni and L. Chittaro, "Stress Detection Using Physiological Sensors," *Computer (Long. Beach. Calif.)*, vol. 48, no. 10, pp. 26–33, Oct. 2015, doi: 10.1109/MC.2015.316.
- [126] C. Kappeler-Setz, "Multimodal emotion and stress recognition," 2012, Accessed: Aug. 27, 2019. [Online]. Available: <https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/63200/eth-5704-02.pdf>.
- [127] Q. Xu, T. L. Nwe, and C. Guan, "Cluster-Based Analysis for Personalized Stress Evaluation Using Physiological Signals," *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 1, pp. 275–281, Jan. 2015, doi: 10.1109/JBHI.2014.2311044.
- [128] A. O. Akmandor and N. K. Jha, "Keep the Stress Away with SoDA: Stress Detection and Alleviation System," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 3, no. 4, pp. 269–282, Oct. 2017, doi: 10.1109/TMSCS.2017.2703613.
- [129] Vanitha V and P. Krishnan, "Real time stress detection system based on EEG signals." Accessed: Jun. 25, 2019. [Online]. Available: www.biomedres.info.
- [130] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. M. Mozos, "Stress Detection Using Wearable Physiological Sensors," Springer, Cham, 2015, pp. 526–532.
- [131] R. Castaldo, L. Montesinos, P. Melillo, S. Massaro, and L. Pecchia, "To What Extent Can We Shorten HRV Analysis in Wearable Sensing? A Case Study on Mental Stress Detection.," Springer, Singapore, 2018, pp. 643–646.

- [132] L. Chen, Y. Zhao, P. Ye, J. Zhang, and J. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Syst. Appl.*, vol. 85, pp. 279–291, Nov. 2017, doi: 10.1016/J.ESWA.2017.01.040.
- [133] A. Ghaderi, J. Frounchi, and A. Farnam, "Machine learning-based signal processing using physiological signals for stress detection," in *2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*, Nov. 2015, pp. 93–98, doi: 10.1109/ICBME.2015.7404123.
- [134] X. Zhang, W. Wen, G. Liu, and H. Hu, "Recognition of public speaking anxiety on the recurrence quantification analysis of GSR signals," in *2016 Sixth International Conference on Information Science and Technology (ICIST)*, May 2016, pp. 533–538, doi: 10.1109/ICIST.2016.7483472.
- [135] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and Galvanic Skin Response signals," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, Jun. 2013, pp. 209–214, doi: 10.1109/CBMS.2013.6627790.
- [136] J. Šalkevicius, R. Damaševičius, R. Maskeliunas, and I. Laukienė, "Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals," *Electronics*, vol. 8, no. 9, p. 1039, Sep. 2019, doi: 10.3390/electronics8091039.

LIST OF PUBLICATIONS

Articles in Journals referenced in the Web of Science:

1. **Šalkevičius, Justas**, Miškinytė, Audronė, & Navickas, Lukas. (2019). Cloud based virtual reality exposure therapy service for public speaking anxiety. *Information*, 10(2:62), 1-15.
2. **Šalkevičius, Justas**, Robertas Damaševičius, Rytis Maskeliunas, & Ilona Laukienė. (2019). Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals. *Electronics*, 8(9), 1039.

Articles referenced in other international databases

1. **Šalkevičius, Justas**, and Navickas, Lukas. (2018). Battling the fear of public speaking: Designing software as a service solution for a virtual reality therapy. *W-FiCloud 2018: 6th International Conference on Future Internet of Things and Cloud Workshops*, 6–8 August 2018, Barcelona, Spain, 209-213.
2. **Šalkevičius, Justas**, and Damaševičius, Robertas. “Adapting Virtual Environments Based on Patients’ Anxiety during Virtual Reality Exposure Therapy.” *11th International Workshop on Data Analysis Methods for Software Systems*, Druskininkai, Lithuania, November 28-30, 2019 (2019): 79.

SL344. 2020-07-22, 12 leidyb. apsk. l. Tiražas 12 egz.

Išleido Kauno technologijos universitetas, K. Donelaičio g. 73, 44249 Kaunas
Spausdino leidyklos „Technologija“ spaustuvė, Studentų g. 54, 51424 Kaunas

