Utrecht University

Ingi Páll Eiríksson
8792283
Department of Information and Computing Science
Human-computer Interaction Master Thesis

Breathing Emotion

Assessing the Respiratory Cues in HCI Emotional Dynamics

First examiner:
Dr. Almila Akdag Salah

Second examiner:
Imke de Jong

Candidate:
Ingi Páll Eiríksson
8792283

Acknowledgements

I am highly grateful for Dr. Almila Akdag Salah, my supervisor, for her guidance and friendly support throughout my research. I also want to thank Imke de Jong for her constructive feedback and patience. I am incredibly thankful for Laís, whose kindness, help and advice have been a great source of strength and confidence. Her generosity and support made a big impact on my thesis and, more broadly, my personal growth. Finally, I feel very fortunate to have chosen Utrecht University in this wonderful city to do my studies. That decision led me to meet an amazing group of dear friends who have enriched my life and made my time there unforgettable. My life in Utrecht and the people I got to know there have been among the highlights of my life, and for that I am deeply grateful.

Abstract

This study looks into the complex dynamics of emotional and physiological responses evoked as individuals engage with various web-based task. The emphasis is on understanding how the emotions elicited by one task can influence both self-evaluated emotion and breathing patterns during the following task. The aim of the research is to show the intricate ways task induced emotions, particularly on the valence and arousal dimensions, effect subsequent user experiences within a digital environment. 19 participants interacted with a controlled environment designed as a clothing store and completed a series of tasks designed evoke different levels of emotions. Throughout these tasks breathing patterns were measured using a respiration belt alongside participants' self-reporting of their emotional states at the end of the study. The findings show evidence that emotions carry over between tasks, particularly in tasks designed to have a greater emotional impact and those that are temporally close to each other. Breathing patterns varied across tasks and showed correlations with self-assessed emotions suggesting a link between the two. The research emphasizes the importance of considering both physiological data and self-reported emotions when understanding web-based interactions. Self-reported arousal showed predictive links across tasks while valence did not. Self-reported valence was however better explained by the breathing patterns over the course of the experiment. This exploratory study offers the field of HCI insights into emotional carry over in web based tasks and how breathing patterns can be used as markers of emotion. It calls for further research to deepen the understanding of the interplay between emotional responses, task characteristics and respiratory markers in a digital environment.

Table of Contents

Terms and Abbreviations

The tables below show terms used for the remainder of the study, their abbreviations and descriptions.

Abbreviation	Meaning	Description
RR	Inhale/exhale breaths per minute	Respiration rate, average number of inhales and exhales per minute.
mDBT	Mean Distance between troughs	Measures the time between peak inhales
mDBP	Mean Distance between breaths	Measures the time between peak exhales
stdDBT	Std Distance between troughs	The variation in time between inhales
stdDBP	Std Distance between breaths	The variation in time between exhales
PH	Mean peak height	Measured depth of breath
PS	Mean peak sharpness	The steepness of the slope following a peak
TS	Mean trough sharpness	The steepness of a slope following a trough
Cc	Character creation	The first task of creating a persona
S	Shopping	The second task of shopping for an item
Ac	Account creation	The third task, a faulty account creation
EoS	Survey	The forth task, the survey at the end
E	Everything	The four tasks combined

Table 0.1. Explanations for the abbreviations to help with the tables in the following sections

Question	Task	Emotionality	Task number
Q3	Character creation	Arousal	1
Q4	Character creation	Valence	1
No question	Shopping	-	2
No question	Shopping	-	2
Q1	Account creation	Arousal	3
Q2	Account creation	Valence	3
Q5	Survey	Valence	4
Q6	Survey	Arousal	4

Table 0.2. A lookup table for the questions and tasks, the shopping task is only defined for the breathing data and has no relevant question.

1. Introduction

In a tale from Germanic mythology, Ondine was an immortal oceanic nymph of incomparable beauty. She had fallen in love with a mortal man. Realizing he was unfaithful she curses him to never draw an unconscious breath again, condemning him to remain awake forever (Demartini et al., 2020; Schestatsky & Fernandes, 2004). This tale finds a chilling echo in reality, manifested as a rare and often fatal disorder known as Ondine's Curse, or Congenital Central Hypoventilation Syndrome. Sometimes caused by a lesion to the lower brainstem and medulla, it forces people to manually control their breathing leading to serious problems when sleeping (Schestatsky & Fernandes, 2004). This parallel between an ancient tale and reality highlights the unique duality of the respiratory process as a vital physiological function, being operated both under conscious control and subconscious automation (Masaoka et al., 2014).

As a primary function, respiration has the vital role of inhaling oxygen into the lungs and throughout the body via the bloodstream followed by the exhalation of carbon dioxide(Sarkar, 2017). Without this simple function, there would not be much more to explore. Breathing is linked to various higher brain functions, it plays an important role in human communication and has a complex and intertwined relationship with emotion (Klausen et al., 2022). People sigh when they are bored or relieved, sob when they are sad, and yawn when they are tired. Breathing rate and depth change with fear, joy, excitement, or stress, reinforcing its intimate connection with our emotions (Kreibig, 2010). Exploring the nuances of this interaction can offer valuable insights and an understanding of human emotion that can prove useful in various ways. A detailed understanding of this bi-directional relationship opens up opportunities for emotion detection systems, improved realism for non-human agents, and can serve as a valuable tool in emotion research. The current research aims to both study and take advantage of this intriguing connection and explore if emotion evoked by a simple HCI task has an impact on performance, perception, or emotional response to a subsequent task, specifically seeking to answer the main research question:

RQ: To what extent does the emotional response evoked by a task influence the emotional response to subsequent tasks, as measured by self-reported emotions and breathing patterns?

The following discussion and research will investigate various theories on how emotions can carry over or morph into something to the detriment or benefit of the following task. Emotional carry over between tasks can be thought of as how much of the emotion still lingers from the previous task but it also encapsulates the physiological responses, such as breathing patterns, that accompany them Understanding these carryover effects is interesting, as they could influence performance, decision-making, and overall experience in subsequent activities. To investigate this the first sub question has to be answered.

SQ1: In what ways do emotional responses, as measured by self-reported emotions and breathing patterns, vary across simple web-based tasks?

Measuring respiration alongside this process can shed light on how breathing and emotion go hand in hand when compared with measures of self-report and might bring insight into the question of whether respiration can become a key to decoding emotion. Self-reporting on emotions has a long history in psychological research and various other fields. Although it appears to be the obvious choice when assessing emotions, self-reporting has some challenges that impact both the reliability and validity of the results. People sometimes show social desirability bias when reporting on their emotions and, even unknowingly, bias their reports. Self-reporting is limited to the emotions that people are aware of, they have to be represented in consciousness (Pekrun, 2016). There are memory inaccuracies in the reporting of, even recent emotions and it only grows with time (Robinson & Clore, 2002). Combining self-report measures with somatic stimuli like physiological data, in this case, breath analysis, could potentially address these issues and provide a more comprehensive and insightful understanding of emotions. This could offer a more objective perspective on emotional experiences, as they are less susceptible to biases and subjective interpretations. By integrating self-reporting with breathing data, researchers can gather complementary data that can validate the reported emotions, leading to the second sub-question:

SQ2: How does the pattern of breathing correspond to self-reported emotions across different web-based tasks?

Combining self-report measures of emotions with somatic stimuli can lead to more reliable and valid results in emotion research. This of course requires a good understanding of the somatic data and how it relates to emotion. To get a feel for the current landscape of knowledge the existing literature was explored. The aim was to discuss and deconstruct the theories and tools relating to the various aspects of task switching and emotion and how respiration can be used to further that knowledge. It explores emotion in the realm of psychology and HCI from various angles and the advantages and disadvantages of different tools of measurement are laid out. The relationship between breathing and emotion is investigated on a biological and a more practical level with a look into the exciting possibilities this offers HCI. The complex cognitive aspects of task-switching and how emotion can alter the process are detailed as how the interplay of emotion, attention, and memory can have a significant effect on tasks and task-switching behaviors. Finally, the nuances of different HCI tasks evoking different emotions and how they can transfer and impact the following tasks are discussed through the lens of the landscape preference model and regulatory focus theory.

In this study, a methodology is proposed to examine the impact of emotional states on task-switching behavior and subsequent emotion in human-computer interaction. By incorporating breath analysis as an objective measurement of emotional states, complemented by self-reported emotion assessment. The aim is to provide an understanding of how induced emotions, such as frustration, influence a user's adaptability and responsiveness in the transition to a new task and how they could vary between individuals. This leads to the final two sub-questions:

SQ3: How predictive are initial task-induced responses, in terms of breathing patterns, and emotional self-assessments of subsequent task responses?

SQ4: What are the implications of the observed relationships for understanding emotional carry-over effects in web-based tasks?

This research intends not only to explore the interplay between emotions and performance but also to reveal potential strategies to better design and manage interactive systems in contexts where task-switching is prevalent. The findings will contribute to enhancing user experience and productivity, and add to the knowledge on how emotion can be incorporated into HCI systems.

2. Literature review

2.1. Breathing and Emotions

2.1.1. Understanding and Interpreting Emotions in HCI

2.1.1.1. The Intuitive Human Ability: Recognizing Emotions

For most humans, recognizing if an individual is experiencing a particular emotion is often an intuitive process. Walking into a room and quickly detecting that a friend is distressed or upset without them saying a word. This intuitive judgment is based on various non-verbal cues such as posture, facial expressions, head movement, eye contact, and a host of other subtle and nuanced cues (Roter et al., 2006). This process is imperfect, and people make biased judgments, but it is central to human communication and social interaction. There is a vast amount of research literature investigating and categorizing these subtle cues and how they are interpreted (Sauter, 2017). Investigating how humans use these subtle cues to communicate emotion is far from being a simple task, something that will turn out to be a common thread when discussing emotions. There are a lot of variances based on context, the level of the relationship between individuals where the accuracy of correctly identifying emotions is affected by the closeness of the relationship (Sternglanz & DePaulo, 2004). There are well-documented individual differences in this emotional recognition, women, for example, tend to recognize some emotions better than men, and the ability to correctly identify emotions seems to decline with age (Abbruzzese et al., 2019).

As is often true when translating complex and intuitive human abilities into a research context, measuring people's emotional states is a great challenge in affective science and human-computer interaction (Mauss & Robinson, 2009). Understanding of human recognition of emotions is far from complete so replicating that process in these contexts requires complex and innovative measures and theories about biological, psychological, and social processes to accurately measure and interpret. This section provides a brief overview of the physiology of emotion, discusses relevant theories, and how they can be relevant to HCI, and highlights some of the challenges of understanding emotion

2.1.1.2. Theoretical Frameworks for Understanding Emotion

A consensual, component model of emotional responding is a helpful representation of the stages of emotional responses. The response starts with an appraisal of a personal significance to a certain situation. This evokes a multifaceted emotional response consisting of subjective experience, a physiological response through the peripheral/autonomic and central nervous systems, and a behavioral response. This emotional response to a situation or a stimulus can be thought of as either having discrete patterns of response for each of the systems or dimensional (Mauss & Robinson, 2009). A good way to explore the qualitative aspects of this is to look at some of the more prominent theoretical frameworks for understanding emotion. The physiological implications of these different views will be explored in a later subsection.

2.1.1.3. The Circumplex Model of Affect: Valence-Arousal Model

One of the most widely adopted frameworks for emotional understanding is the circumplex model of affect (Posner et al., 2005) often referred to as the valence-arousal model. The model was developed by Russel in 1980 and is a dimensional model that quantifies emotions as points on a two-dimensional plane with the center representing a neutral state. The dimension usually represented on the horizontal axis is valence, ranging from negative to positive or the pleasantness or unpleasantness of an emotion. Valence contrasts emotions like happiness and sadness on the opposite sides of its spectrum. On the vertical axes lies arousal ranging from high to low or the activation or deactivation of the emotion. Here emotional states like sleepy and surprised could represent the extremes on the spectrum (Basu et al., 2015; Mauss & Robinson, 2009; Nandy et al., 2023). Having originally been developed in 1980 the valence-arousal model has proved to be a successful method of understanding and quantifying emotions and is frequently used in research 40 years after its creation (Klausen et al., 2022; Roes et al., 2022a). The model is also the foundation behind some of the most widely used emotional scales and measuring tools. The Self-Assessment Manikin (SAM) is one of the most popular self-assessment scales of affect and is based on a modified version of the model (Betella & Verschure, 2016; Bradley & Lang, 1994).

There are more dimensional frameworks of emotions that agree with the general idea that there are limited underlying dimensions that emotions can be organized. They however differ in some specifics like the number of dimensions. The valence-arousal has had some criticism. For example, there is some amount of overlap in emotions. Fear and anger can occupy a highly similar space on the two-dimensional plane with both being high arousal and negative affect (Mauss & Robinson, 2009). This has been addressed by introducing a third dimension to the model. This third dimension is frequently called dominance representing the degree of control generated by the stimulus (Basu et al., 2015). This approach was for example used in the development of SAM (Betella & Verschure, 2016). This third dimension is sometimes represented as an approach-avoidance dimension that traditionally was thought to be explained by valence. Emotions like anger and fear, however, suggest there is a dissociation between those dimension. They share a similar profile on the valence/arousal dimensions but are opposed to each other on the dominance dimension with fear causing withdrawal and anger inciting engagement (Mauss & Robinson, 2009). Another question regarding the valence-arousal model is the relations between the two dimensions. Research suggests that the three dimensions are not independent of each other and some point out a “V-shaped” association between the two so when either negative or positive valence intensifies, arousal also escalates(Kuppens et al., 2017; Nandy et al., 2023).

The valence-arousal model is used in HCI research and application (Klausen et al., 2022). The model is a valuable asset due to its simplicity and ability to quantify emotions, making them useful for empirical research and implementation into systems through emotion recognition. Both dimensions, although arousal is often considered more universal and easier to measure (Critchley & Nagai, 2013; Lim, 2016), are associated with measurable physiological changes that can have applications for HCI (Kreibig, 2010). Being able to use physiological markers to automatically assess emotions has various use cases for adaptive systems and beyond. This will be discussed in more detail in section 2.1.3 “*Emotions and Breathing Patterns as Emotional Indicators”* . The subjective nature of valence and the variations due to contextual factors can be problematic for HCI research and can result in some inaccuracies. Two individuals can show similar physiological responses in terms of arousal but interpret them differently due to various individual differences and contexts (Kreibig, 2010; Lim, 2016). This suggests it can be detrimental to simply rely on emotion-specific response profiles for emotion recognition (Pace-Schott et al., 2019).

2.1.1.4. Basic Emotion Theory: A Discrete Perspective on Emotion

In contrast to the dimensional way of understanding emotions as exemplified by the valence-arousal model, there are theories of discrete emotions. Emotions should be thought of as distinct and fundamental categories each with a unique profile of physiology, subjective experience, and behavior (Mauss & Robinson, 2009). One of the most influential theories of this view is the basic emotion theory, proposed by Ekman. Ekman identified six basic emotions: happiness, sadness, fear, anger, surprise, and disgust(Basu et al., 2015; Ekman, 1992). Ekman argued that emotions were genetically determined and expressions of discrete emotions are interpreted the same way independent of culture. The same goes for emotional responses to similar situations (Lim, 2016). There is an abundance of evidence that suggests that people around the world express and similarly recognize emotion supporting the idea that emotion is not learned but genetically encoded (Pace-Schott et al., 2019).

This theory has an intuitive appeal and is supported by empirical evidence that people can generally recognize these basic emotions within themselves and others suggesting they are fundamental human experiences (Pace-Schott et al., 2019). Having strictly classified and universal emotions can be beneficial for HCI. Machine learning algorithms can be trained to classify these basic emotions based on facial expressions or other physiological markers. Having these six universal emotions with their clearly defined responses would be helpful when designing systems or agents that are supposed to express emotion, although it would be rather low resolution. Unsurprisingly, this is not this simple. Although there is evidence that suggests these emotions are recognized across cultures they are not expressed, treated, or accepted the same everywhere. Emotions of high arousal for example are more likely to be promoted and expressed in some cultures compared to others (Lim, 2016). The assumption that each emotion is associated with a unique physiological signature has proven to be difficult to realize. There is a significant overlap in a large range of physiological markers between emotions (Kreibig, 2010). This does not mean that it is impossible to create a unique signature for every given emotion but it would have to take into account various external variables like context and include a wide variety of different measurements of physiological markers.

Summary. The intuitive human ability to recognize and interpret emotions is a multi-faceted process that relies on various non-verbal cues, from facial expressions to posture. This understanding of emotions, while not flawless, plays an integral role in our social interactions and is underpinned by theories such as the circumplex model of affect and the basic emotion theory. The circumplex model, often referred to as the valence-arousal model, uses two dimensions, valence, and arousal, to quantify emotions, offering a simple, effective method of understanding and measuring emotional states. Although it has faced criticism, this model remains widely utilized in HCI research due to its ability to associate measurable physiological changes with emotions, thus enabling the design of adaptive systems that respond to emotional cues. The basic emotion theory posits that emotions are distinct, fundamental categories, each associated with a unique profile of physiology, subjective experience, and behavior. While this perspective provides a more detailed categorization of emotions, proving to be helpful in training machine learning algorithms and designing emotionally expressive systems, it grapples with the challenge of significant overlap in physiological markers across emotions and variations in emotional expression and acceptance across different cultures. Both theories, despite their challenges, offer valuable insights into the complex landscape of emotions and have significant implications for the field of HCI, underscoring the need for a nuanced, context-aware approach to interpreting and integrating emotional cues into our digital interactions.

2.1.2. The Interplay Between Emotions and Respiration

2.1.2.1. The Physiology and Evolutionary Role of Emotion

Emotions and respiration are intricately connected. Rapid breathing in moments of anger or arousal, sighing in times of sadness, or yawning during periods of boredom are all familiar experiences (Boiten et al., 1994). Both emotion and respiration are a part of the autonomous nervous system (Pace-Schott et al., 2019; Roes et al., 2022b). Emotions are more than just a cognitive phenomenon, they encompass a wide variety of bodily responses ranging from autonomic to behavioral responses. Emotions influence blood pressure, pupil dilation, heart rate, skin conductance, and, most importantly for the current research, respiration (Masaoka et al., 2014; Pace-Schott et al., 2019).

Emotions have an evolutionary purpose, serving as adaptive responses to environmental challenges and facilitating survival and reproductive success. They enable animals to quickly and effectively respond to situations, make decisions, and prioritize their actions (Al-Shawaf et al., 2016). Emotions also help with intra-species communication of intentions. They can be understood relatively universally across cultures, infants express emotions like adults and even chimpanzees mirror some of the emotional facial features that humans show (Pace-Schott et al., 2019). Respiration plays a crucial role in the social context as it is connected to emotional regulation and expression (Roes et al., 2022b; Sarkar, 2017). Respiration is a key factor in the physiology of emotional responses. In an emotionally charged situation breathing has to adapt to the demands of the environment. In the presence of a threat, the response can be anger, an emotion that prepares for quick action and requires oxygen to the muscles, triggering rapid breathing (Pace-Schott et al., 2019; Roes et al., 2022b). Conversely, calmness and happiness would cause slower, deeper breathing to promote relaxation and preservation of resources (Roes et al., 2022b; Sarkar, 2017).

2.1.2.2. Neurological Connection Between Emotion and Respiration

The relationship between emotion and respiration continues to be intimate on a neurological level. There are, speaking in low-resolution, simple terms, three main brain areas that regulate respiration: the brainstem, the limbic system, and the cerebral cortex (Masaoka et al., 2014). The primary function of respiration is the process of inhalation and exhalation to carry oxygen into the lungs and carbon dioxide out. This function comes effortlessly and without thought (Del Negro et al., 2018; Sarkar, 2017). Breathing is by default an unconscious activity in the medulla oblongata, a primal brain area deep In the brainstem that concerns metabolism, maintains homeostasis, and is vital for other essential life functions(Del Negro et al., 2018; Sarkar, 2017). More importantly for the current research, the limbic system, which is involved in the processing and regulation of emotions(Masaoka et al., 2014), plays a role in controlling breathing through its connections with the respiratory centers in the brainstem. This has, for example, been shown in animal research where stimulating the amygdala increases respiratory rate and induces anxiety (Homma & Masaoka, 2008).

Something that sets respiration aside from other vital autonomous bodily functions is the fact that it can be altered by the cerebral cortex, particularly the motor cortex, through voluntary actions (Homma & Masaoka, 2008; Masaoka et al., 2014). This top-down control over respiration makes it fundamentally different from other vital functions like heart rate for example. But this control serves as a way to regulate emotion and thereby other physiological functions that follow such as heart rate and blood pressure (Masaoka et al., 2014; Pace-Schott et al., 2019). *"The brain, by regulating breathing, controls its own excitability" (Balestrino & Somjen, 1988)*. In summary, the rhythmic breathing "signal" is sent from the brainstem, altered by the limbic system to fit the appropriate emotions while being impacted by a complex bi-directional interaction with the brainstem to produce the final respiratory "output" (Homma & Masaoka, 2008).

Summary. Emotions and respiration are intertwined and fundamental components of the autonomous nervous system. They are adaptive responses to environmental challenges, facilitating survival, decision-making, and communication. The pattern of respiration can change depending on the emotional situation, preparing the body for action or promoting relaxation. The two are similarly intimately linked at the neurological level, with the brainstem, limbic system, and cerebral cortex playing crucial roles in the regulation of respiration. This connection allows the brain to control its own excitability by regulating breathing, making respiration distinct as a vital function that can be consciously altered[1].

2.1.3. Emotions and Breathing Patterns as Emotional Indicators

2.1.3.1. Measuring Emotions in HCI

Accurately measuring, and correctly evoking emotions is an important aspect of emotional research in HCI and every field. The complexity of emotions and their intertwined relationship with cognitive and physiological mechanisms make this a difficult task that requires sophisticated methods (Masaoka et al., 2014; Pace-Schott et al., 2019). This is not simplified by the fact that emotions can be unconscious and difficult to realize (Pekrun, 2016). Traditionally, self-reporting tools such as

[1] The summaries in this paper are generated by OpenAI's ChatGPT based on the content of the subsection and modified as needed.

questionnaires and interviews have been used to assess the emotional states of individuals. These methods have the advantage of, not only measuring the current emotional state but give insight into past emotional states. Self-report of emotion is often the only method available and offers a more nuanced insight into emotions and related thoughts (Pekrun, 2016). This method is not without its limitations as has been discussed in some detail in the *Challenges and Solutions in Emotion Research* subsection.

As the understanding of the physiological aspects of emotions grows, more objective approaches to measuring emotions become available. These methods rely on physiological markers and somatic stimuli to offer a different perspective. These methods include heart rate monitoring, skin conductance, and respiration pattern analysis. The use of breathing patterns as a physiological measure of emotions offers a promising method of emotion detection but is not without its challenges. Breathing patterns are impacted by emotional states, varying in rate, depth, and regularity (Roes et al., 2022b). Monitoring these parameters can offer a different type of insight into emotional responses and adds a point to the nomological network of emotion recognition. Analyzing respiration patterns for this purpose comes with its challenges as they are influenced by a variety of non-emotional factors (Pessanha et al., 2022).

To measure emotions in a research setting they must first be evoked in some manner. This requires carefully chosen and controlled stimuli that can reliably trigger a range of emotions. This brings a multitude of possible problems such as individual differences in responses to the stimuli and the complexity of finding material that can evoke a specific emotion. The following section will further detail how emotions are measured and evoked in HCI research. Different methods of self-reporting will be discussed in some detail with their advantages and disadvantages. The use of different physiological markers will be explored with a particular focus on the potential of utilizing breathing patterns as emotional indicators.

2.1.3.2. Self-Reporting Scales for Emotional Assessment

The advantages and limitations of self-reported emotions have been addressed, but not all self-report measures are created equal. Various scales have been developed to systematically assess subjective emotion and each has its unique attributes and potential drawbacks (Mauss & Robinson, 2009; Pekrun, 2016). As with any measurement tool, to be successful, emotional assessment scales have to be both valid and reliable, that is they have to consistently measure what they are supposed to measure and highly correlate with other established measures of the same construct. This requires them to have construct specificity, each dimension on the scale should be measuring a single mental construct. This can be troublesome when measuring emotions due to the components of different emotions overlapping. This construct overlap is further increased by the often unclear language used when discussing emotions. These factors make discriminant validity a problem worth focusing on when choosing a scale (Pekrun, 2016). Emotions are multifaceted so any given scale is unlikely to capture the whole spectrum. Making sure that the chosen scale measures the appropriate emotions and that their definitions align with the research is important. Finally, a good scale should be versatile across ages and cultures, and easy to use as this is one of the strengths of self-reporting measures (Pekrun, 2016).

The Self-Assessment Manikin (SAM) as proposed in 1994 is still one of the most popular self-assessment scales of affect (Betella & Verschure, 2016; Bradley & Lang, 1994). SAM is a non-verbal picture-oriented scale that uses five images for each of its three dimensions that users rate on a scale of 9 or 21 points. It can be used to measure an emotional response to a wide variety of stimuli (Bynion & Feldner, 2017). The three dimensions are; pleasure, ranging from unhappy to happy, arousal, ranging from relaxed to excited or stimulated, and dominance, ranging from feeling controlled to feeling in control. In practice they range from a frowning to a happy figure, a sleepy to an alert or wide awake figure, and a small figure to a large figure respectively (Betella & Verschure, 2016).

SAM has been shown to have good psychometric properties with acceptable reliability and validity across various tests (Backs et al., 2005; Nabizadeh Chianeh et al., 2012) making it a trustworthy tool for measuring these dimensions. SAM has more advantages that likely explain some of its popularity. The scale is short and easy to use which allows for multiple measurements in a single session (Bynion & Feldner, 2017). Its simplicity makes it usable for children as well as adults, and the scale has been shown to have acceptable psychometrics for both age groups (Backs et al., 2005). The fact that it does not rely on language allows for its use for young children and across different languages and cultures. Finally, SAM is versatile as it can be used to measure emotional response, as it was designed to do, but also how a person feels "in general" and has even been used to determine how another individual feels (Klausen et al., 2022). It should be noted that the evidence for its psychometric properties should not be assumed to be as robust for those use cases.

SAM has been criticized for its dated appearance and its paper-and-pencil design is no longer intuitive with theories of modern graphical interfaces being different (Betella & Verschure, 2016). SAM also only captures three dimensions of human emotion and therefore leaves out a lot of nuances. That is something that can be expected and can be viewed as quality over quantity. Overall, SAM is a relatively sound and simple-to-use scale that mitigates some of the problems with self-reporting of emotion.

Even with the shortcomings of self-reporting useful methods to gather data on emotional states exists. These tools should be used with care and the nuances of their interpretation should be respected as it can be easy to jump to conclusions when dealing with something as subjective as emotional self-reporting. The next subsection will discuss physiological markers of emotion and how they can be used to supplement self-reporting for a more robust understanding.

Summary. The importance of accurately measuring and evoking emotions in HCI and other fields is emphasized, underlining the complexity of the task due to the intertwined relationship of emotions with cognitive and physiological mechanisms. Traditional methods such as self-reporting tools, including questionnaires and interviews, offer valuable insights into both current and past emotional states. However, they also pose limitations, as discussed in a previous section. As understanding of the physiological aspects of emotions advances, more objective measurement methods relying on markers like heart rate, skin conductance, and respiration patterns are emerging. These methods offer new perspectives but also bring their challenges.

To measure emotions in research, they must first be effectively evoked, which entails the careful selection and control of reliable stimuli. The following section delves into different self-reporting methods and their pros and cons, with a special focus on the use of physiological markers, particularly breathing patterns, as indicators of emotion. The subsequent subsection reviews self-reporting scales for emotional assessment, noting the importance of validity, reliability, and construct specificity. The Self-Assessment Manikin (SAM), a popular non-verbal picture-oriented scale, is discussed in detail, outlining its advantages such as simplicity, versatility, and psychometric properties. However, criticisms such as its dated appearance and its limitation to only three dimensions of human emotion are also mentioned. The need for careful use and interpretation of self-reporting tools is stressed, highlighting their inherent subjectivity. The final subsection promises to explore physiological markers of emotion, complementing self-reporting methods for a more comprehensive understanding of emotions.

2.1.3.3. Inducing basic emotions

The most robust method of testing the causal effect of emotions on physiological and psychological variables is directly inducing them during an experiment (Siedlecka & Denson, 2019). When the elusive definitions of emotions, and the competing theories about their nature discussed in the previous sections are taken into account it is clear that this task requires nuance and precision. In addition to the type intensity of the stimulus, the influence of emotions depends on personality differences as well as situational context (Kučera & Haviger, 2012). There are various commonly used methods for inducing emotions. They are not all equally effective and differ in their ability in evoking specific emotions (Siedlecka & Denson, 2019). As it is important to be careful in picking the appropriate assessment scale or physiological measurement, the stimulus should be carefully chosen to align with the research goals and the emotions under investigation. It is important to keep in mind that a specific stimulus can evoke a range of emotions, some of which will be unexpected the researcher making noise in the physiological data and giving unexpected results in the self-assessment questionnaires. Take for example the intention to evoke happiness by having a subject eating chocolate with the task unexpectantly inducing guilt as well (Siedlecka & Denson, 2019).

The stimuli used to evoke emotions should ideally be empirically evaluated before the experiment. It should be noted, as previously discussed, it could prove problematic to simply use self-report or physiological measurements for the evaluation. Studies have shown that physiological measurements can be contradictory within a single emotion and that they sometimes differ on one dimension even if they agree on the other (Siedlecka & Denson, 2019). When evaluating the stimulus it is preferable to use multiple different measurements to increase construct validity. This subsection will briefly discuss different methods of invoking emotion, their limitations, advantages, and interactions with some of the classic emotions. The interactions with the emotions are based on Siedlecka & Denson’s 2019 evaluation of five common methods and the effect on various emotions based on, both, subjective experience and physiological responses (Siedlecka & Denson, 2019).

A frequently used method is displaying material selected for emotional impact. This can be visual stimuli like pictures or video or auditory stimuli (Kučera & Haviger, 2012). Visual stimuli were shown to be the most effective inducing method for all the tested emotions: anger, happiness, fear, disgust, surprise, and sadness. This method has the advantage of having a large corpus of available labeled stimuli like the International Affective Picture System (Siedlecka & Denson, 2019). However, responses can be influenced by individual differences in interpretation and experiences. The same is true for audio stimuli where musical taste differs between individuals and cultures. Audio is however a powerful tool to induce emotions and was found to be primarily good for inducing happiness, fear, and sadness. The use of imaging techniques to induce emotions, like autobiographical recall techniques can induce strong emotions (Kučera & Haviger, 2012) in a personalized context. This method was found to be effective at inducing all tested emotions besides, somewhat predictably, surprise (Siedlecka & Denson, 2019). Researchers should be wary of ethical connotations when using autobiographical recall to induce negative emotions. Finally, preset interactions or situational procedures. These tasks can be games, social situations, or individual tasks designed to provoke certain emotions. They can be diverse and highly engaging. When tested as social situations they were found to be effective in inducing anger, surprise, fear, and happiness (Kučera & Haviger, 2012; Siedlecka & Denson, 2019). These tasks or situations can be complex to design and standardize and can be susceptible to individual differences in interpretation and performance which might impact emotional responses.

It is worth mentioning that a combination of methods could be more effective than a single paradigm. Autobiographical recall has been shown to have a greater effect when paired with exercise and music has been known to increase the emotional impact of a movie or a video game. This should not be taken as a certain and should be evaluated on a case-by-case basis (Siedlecka & Denson, 2019).

Summary. This subsection has elucidated the importance of tactfully and accurately inducing emotions in an experimental context, considering the various theories and elusive definitions of emotions. It underlines that apart from the stimulus's type and intensity, emotional responses may also hinge on individual personality variations and the situational context. Various techniques for emotion induction are employed, each offering different levels of effectiveness in eliciting specific emotions. Therefore, a careful selection of stimuli, aligning with the research objectives and the emotions under investigation, is paramount. This discussion includes an examination of several methods of emotion induction such as visual and auditory stimuli, imagination techniques, and preset interactions or situational tasks. Each method presents unique advantages, limitations, and ethical considerations, while also exhibiting differential effectiveness across emotions.

2.1.3.4. Physiological Measure of Emotion

Measuring the physiological aspects of emotion offers the possibility to quantify and analyze the physical changes that occur in response to an emotional stimulus. This most generally involves measuring bodily function controlled by, or influenced by, the autonomic nervous system (ANS). The ANS is associated with excitation and inhibition through its two branches, the sympathetic and parasympathetic nervous systems (Appelhans & Luecken, 2006; Mauss & Robinson, 2009). The fact that the ANS does not exclusively function as an emotional response system complicates this process with the ANS being involved in other functions such as maintaining homeostasis and digestion. This means that the variability in the measurements cannot be attributed solely to the emotion but could instead be caused by other functionality of the ANS (Mauss & Robinson, 2009).

A debate that originated in the late 1800s with the James-Lange theory of emotion is still central to research on physiological markers of emotion to this day (Coleman & Snarey, 2011; Pace-Schott et al., 2019). Emotions and their physiological manifestations can be thought of as discrete or dimensional. The idea of discrete emotions is an extension of James-Lange theory and explains, in simple terms, that the specific feelings that are elicited by physiology stem from emotions having specific and unique patterns of ANS responses. More specifically, emotions are simply experienced because of the physiological changes followed by a stimulus (Pace-Schott et al., 2019). It is obvious to see how this theory would emphasize the importance of the measurement of physiological parameters for emotional recognition. There is good evidence for discrete emotions having some ANS specificity. Emotional classification using machine learning has reached impressive accuracy. One study achieved an 83.7% recognition rate classifying six discrete emotions: anger, sadness, amusement, frustration surprise, and fear. Another one reached 92.05% between the four emotions anger, joy sadness, and pleasure. These studies used multiple physiological markers including electrocardiogram, electromyography, skin conductivity, and respiration signals (Basu et al., 2015; Pace-Schott et al., 2019).

The dimensional side of the debate organizes emotions by some underlying constructs, such as valence and arousal (Mauss & Robinson, 2009). In this case, emotions do not have ANS specificity or unique patterns but map onto broad dimensions. Here the important role of the ANS in emotion is mediated by top-down appraisals (Pace-Schott et al., 2019). There is good evidence for a more continuous model of emotion and the respective physiological markers. An example study reached 89.9% and 96.6% accuracy for valence and arousal using a neural network classifier and signals from skin conductivity, blood volume pulse, electrocardiogram, respiration, and electromyography (Basu et al., 2015). The implications of thinking about physiological measurements continuously instead of categorically change the approach when trying to assess emotional states. As with many complex things both theories have something to offer and can complement each other. Research suggests that both play a role, indicating a continuum of autonomic specificity with various levels of differentiation within the ANS. The situation is complicated by the fact that the extent of ANS specificity seems to depend on several variables, including the context, individual variances, and how the emotion is evoked (Pace-Schott et al., 2019). A combination of the two theories can be thought of as describing what appears to be a "discrete" emotion by a combination of levels on multiple dimensions, fear for example could be high arousal, negative valence, and low approach motivation (Mauss & Robinson, 2009).

These differing theories underline the fact that physiological markers should not be taken at face value and there is debate surrounding how they should be interpreted to assess emotion. They do however complement self-reporting methods by adding objective, quantitative, and often real-time data that can be analyzed and used to compare individuals and emotional states (Pace-Schott et al., 2019). There are different methods and tools used to measure these diverse signals. They differ in their intrusiveness, complexity of use and analysis, and the intended purpose. Some of these methods will be discussed briefly below.

There are various ways to use cardiovascular parameters to estimate ANS activation and emotional reactions. Some of the most common are heart rate, blood pressure, and heart rate variability (HRV) (Mauss & Robinson, 2009). These have the advantage of being relatively easy to measure by fairly non-invasive means such as heart rate monitors. They can also offer data in real-time and stemming from the ANS they cannot be hidden or faked by a subject (Basu et al., 2015). An expected complication using cardiovascular markers is the large amount of non-emotional effects that can impact the variance in the measurements, physical health, or activity. HRV is especially interesting as it is the only psychophysiological variable that offers information on both branches of the ANS. It can therefore be used as a marker for both inhibitory and excitatory processes in emotion regulation and responding with promising implications for mental health research (Appelhans & Luecken, 2006).

A frequently used marker is skin conductance or electrodermal activity. This method measures the skin's ability to conduct electricity through minuscule changes in sweat production (Mauss & Robinson, 2009). Skin conductance level is commonly used to measure arousal as changes in sweat production are a physiological response to excitement or stress, it is however used to measure a broader range of emotions. Measuring skin conductance is inexpensive and relatively non-invasive. It can however be difficult to interpret as it is sensitive to a broad range of stimuli. It is a good marker for arousal but is not a good indicator for different types of arousal (Critchley & Nagai, 2013).

Finally, different measures are used to measure emotions by directly looking at brain activity as it can reflect different emotional states. Functional Magnetic Resonance Imaging (fMRI) measures brain activity by measuring blow flow between brain regions and offers the best spatial resolution in emotion research. fMRI, however, has a low spatial resolution that creates problems when dealing with the dynamic and responsive nature of emotions (Pace-Schott et al., 2019). FMRI is highly invasive, difficult to use, and expensive. A solution to the temporal resolution problem is Electroencephalography (EEG) which can give real-time information on brain activity triggered immediately after stimulus presentation. EEG is, however, invasive and complex (Pace-Schott et al., 2019). Different measurements of physiological markers of emotion have been briefly discussed here, in the next session respiration patterns will be discussed in more detail.

Summary. This section discusses the measurement of physiological aspects of emotions, highlighting the role of the autonomic nervous system (ANS) in controlling bodily responses to emotional stimuli. The complexity of these measurements is noted due to the multifunctionality of the ANS, which also regulates homeostasis and digestion. Two contrasting theories of emotions' physiological manifestations, discrete and dimensional, are discussed. The discrete theory, based on the James-Lange theory, proposes unique ANS responses for each emotion, while the dimensional theory organizes emotions according to underlying constructs like valence and arousal. Both theories have empirical support from studies using machine learning and neural network classifiers to recognize emotions based on physiological markers. The debate between these two theories emphasizes the complexity of interpreting physiological markers for emotional assessment. Despite this, they complement self-reporting methods by providing objective, quantitative, and real-time data. Several methods to measure these signals are discussed, including cardiovascular parameters such as heart rate, blood pressure, heart rate variability, skin conductance or electrodermal activity, and neuroimaging techniques like fMRI and EEG. Each method has its advantages and disadvantages in terms of intrusiveness, complexity of use, and analysis. The ensuing section will delve deeper into the specifics of respiration patterns as a physiological marker of emotions.

2.1.3.5. Breathing Patterns as a Physiological Measure of Emotion

The intimate and bidirectional connection between respiration and emotion has been discussed in some detail. Following the previous subsection, it is natural to highlight the utilization of breathing patterns as a physiological measure of emotion. It can be expected that some of the limitations of other somatic measures apply to respiration as well. Knowing that breathing is a product of several different factors, some of which are not related to emotions suggests this is a complex task (Masaoka et al., 2014). As with other markers, various parameters can be measured, and different methods and gadgets to measure them. These methods have their advantages and challenges, they differ in their intrusiveness, complexity of use and interpretation, and with the intended purpose. Some of these parameters and methods will be discussed here with examples of how they have been associated with discrete emotional states in the literature.

Breathing can be measured in terms of several quantifiable parameters that have been shown to relate to emotion. Unsurprisingly, many of those measure either frequency or volume. Respiratory rate (RR) is a common and straightforward measure, simply measuring the number of breath cycles per minute. When interpreting respiration rate it is important to keep in mind that breathing is not stable, there is a breath-to-breath variation that can be significant even though the RR stays the same (Drummond et al., 2020). Changes in RR can indicate a change in emotional state, with higher RR associated with emotional arousal whether it is excitement due to happiness, fear, or anger (Kreibig, 2010; Siddiqui et al., 2021). Inspiratory (Ti) and expiratory (Te) times refer to the duration of the inhalation and exhalation phases of the breathing cycle, often including expiratory pauses (Gomez & Danuser, 2004). There is evidence for both variables, and the ratio between them being used as markers for various emotional states. Fear has been associated with a higher Ti/Te ratio and disgust with a decrease in Ti (Kreibig, 2010). The amount of air displaced during a normal breath is called tidal volume (Vt), and is either measured as inspiratory volume or expiratory volume (Gomez & Danuser, 2004) Disgust has been connected with a decreased Vt (Kreibig, 2010). Minute ventilation is the total volume of air entering or leaving the lungs every minute (Gomez & Danuser, 2004). Anxiety is associated with decreased minute ventilation and higher RR through a decreased Ti and increased Te. Amusement with increased RR, increased respiratory irregularity lower Ti, Vt and Ti/Ttotal, meaning the inhalation is shorter than the exhalation. (Kreibig, 2010). These are just examples of common measurements. A more unconventional measure is respiratory sinus arrhythmia. Heart rate variability and its role in emotion regulation were briefly discussed in the last section, the frequency of heart rate variability is in some part due to respiration as heart rate naturally increases with inspiration and decreases with expiration, this is called respiratory sinus arrhythmia (Appelhans & Luecken, 2006; Siddiqui et al., 2021). As discussed about other physiological measurements of emotion there is more complexity to the aforementioned examples of respiratory representations of discrete emotions. They are like others dependent on the context, individual variances, and how the emotion is evoked (Pace-Schott et al., 2019). To name an example, the different stimuli used to evoke disgust appears to be important when measuring breathing patterns, with injury-related disgust and contamination-related disgust having different response. The same has been found with happiness evoked by visual stimuli showing different patterns than for other emotion-evoking paradigms (Kreibig, 2010). This is something worth further research.

2.1.4. Practical Implications of the Emotion-Respiration Relationship

2.1.4.1. Challenges and Solutions in Emotion Research

Having briefly reviewed the intimate relationship between emotions and respiration, it is worth discussing the practical applications this understanding brings to various fields. This relationship has implications for research and application in multiple disciplines including psychology and HCI. Self-reporting on emotions is a method used for emotion research in HCI and in various other disciplines. Although it appears to be the obvious choice when assessing emotions self-reporting has some challenges that impact both reliability and validity of the results (Goetz et al., 2013; Pekrun, 2016; Roes et al., 2022b). People sometimes show social desirability bias when reporting on their emotions and, even unknowingly, bias their reports. Self-reporting is limited to the emotions that people are aware of, they have to be represented in consciousness (Pekrun, 2016). There are memory inaccuracies in the reporting of, even recent emotions and it only grows with time (Robinson & Clore, 2002). Recall of memories can be influenced by the current emotional state which can cause challenges when researching changes in emotions (Kensinger, 2009). Finally, people differ in their introspective abilities and their ability to articulate emotions where individual differences can be significant.

Combining self-report measures with somatic stimuli like physiological data, in this case, breath analysis, could potentially address these issues and provide a more comprehensive and insightful understanding of emotions. This could offer a more objective perspective on emotional experiences, as they are less susceptible to biases and subjective interpretations. By integrating self-reporting with breathing data, researchers can gather complementary data that can validate the reported emotions. Combining self-report measures of emotions with somatic stimuli can lead to more reliable and valid results in emotion research. This of course requires a good understanding of the somatic data and how it relates to emotion.

The ability to control respiration and emotion top-down has been taken advantage of for a long time and in a multitude of ways. Controlled breathing has been used in yoga practices for a long time and more recently in the medical literature where it has been used to control panic attacks and promote relaxation (Roes et al., 2022b). Breathing exercises have been used in combination with cognitive behavioral therapy to combat depression amongst other problems (Chien et al., 2015).

2.1.4.2. Applications in HCI and Affective Computing

Focusing on the intersection of emotions, breathing, and HCI research, there are many interesting applications and use cases where this knowledge is useful. Affective computing deals with systems and devices that can recognize, interpret, process, and simulate human emotions. These systems could use breathing patterns as one of many biometric inputs to determine a user's emotional state. For example, a system might detect rapid, shallow breathing as a sign of stress and adapt accordingly. Systems that can automatically assess emotional state from voice and tools that use motor activity recordings to detect depression are being actively researched (Frogner et al., 2019; Pessanha et al., 2022). Similar machine-learning techniques could be applied to breathing data. There is a high correlation between respiration patterns and emotion, although it can depend on context (Siddiqui et al., 2021). But as previously discussed breathing patterns are modulated by multiple factors. That creates the complication of distinguishing the emotion-related factors from the others for them to prove useful in analyzing emotion or mood disorders (Pessanha et al., 2022).

The benefits of determining emotions based on physiological signals have been discussed. There is a caveat that this often involves intrusive methods of measuring those signals. Different measures require connecting wires to various body parts or putting on a belt or a helmet (Siddiqui et al., 2021). Although these methods are usable in a research environment there is a clear advantage of using less intrusive methods. Respiration pattern analysis could be a prime candidate as a non-invasive measurement of emotionality. In a 2021 study, researchers used a non-invasive radio ultra-wideband radar to gather respiration data from participants. They focused on happiness, fear, and disgust and induced emotions with movie clips while the radar measured chest movements. The data was then used to train and test a machine-learning algorithm. The method had the potential to differentiate between the emotions with 76% accuracy and additionally shed some light on the different ways in which male and female participants expressed those emotions (Siddiqui et al., 2021). This study shows the potential of using respiration as a means to recognize emotions even if it only focused on three emotions. With the ever-advancing capabilities of machine learning algorithms, this is sure to be an interesting possibility.

2.1.4.3. Enhancing Realism in Virtual Agents

Having a good understanding of the relationship between breathing and emotions can have a significant impact on the development of virtual agents. It is said that as virtual agents become more realistic, the need for their models to be realistic at a detailed level grows (Bernardet et al., 2017).

While virtual agents of various kinds can speak, adding the nuance of dynamic breathing, sighs, sobs, and other respiratory behaviors could add to their realism, especially when portraying emotions (Bernardet et al., 2017; Roes et al., 2022b). The implementation of appropriate respiration behavior could add to the user experience in a variety of ways and levels of virtual agents. An automatic screen reader could be made to sound more human-like by adding respiration and emotionally relevant breathing. Animated agents could have more realistic emotions if bodily respiratory behavior is displayed, displaying rapid breathing in a state of fear for example. By mirroring users' breathing patterns, as well as other non-verbal behaviors, virtual, or robot, agents could exhibit a form of empathy, further enhancing their realism and the feeling of empathy. Mimicking user's non-verbal emotional expressions related to perceived empathy in simulated consultations (Zhou & Fischer, 2018) although breathing patterns specifically were not involved in the study.

A 2022 study aimed to investigate how humans attribute emotions to a non-human soft robot. The robot displayed an alternating frequency of inflating and deflating which the researchers referred to as breathing. The focus of the experiment was to address the feasibility of using breathing rates as a non-verbal means for a robot to communicate distinct emotions. The emotions were measured on a validated scale usually used to self-report pleasure, arousal, and dominance. In this case, It was used to assess the emotionality of the robot. They found that participants could easily recognize pleasure and arousal based on the breathing rates of the robot, even if breathing rate was the only cue they had. Dominance was harder to determine by breathing rates and there was no significant difference between different rates (Klausen et al., 2022). This study supports the idea that breathing patterns can be used to enhance the interpretation of emotion in virtual agents, at least non-human agents.

The benefits of integrating simulated respiration into virtual agents are not undisputed. Novick et al.(2018), after interacting with SimSensei, a chat agent developed for diagnosing PTSD, observed that it did not exhibit any breathing patterns. This observation led to a 2018 study in which they created an agent with simulated breathing, known as Paola Chat. The study included an empirical evaluation of the human subject's perception of the agent's breathing amplitude to optimize its realism. When evaluating Paola Chat, the researchers found no significant differences in terms of perceived naturalness, rapport, or presence when compared to an agent without animated respiration. This study is not without its limitations, suggesting the need for further investigation. Specifically, while Paola Chat exhibited animated breathing, it did not include any breathing sounds. The researchers' initial observation of SimSensei's lack of breathing might have predisposed them to assume that the inclusion of breathing would enhance the agent's naturalness. A noteworthy difference between SimSensei and Paola Chat was their level of communication. SimSensei was largely silent during interactions, whereas Paola Chat was actively conversing. This might imply that the perceived importance of animated breathing could increase when the agent is not actively communicating (Novick et al., 2018).

Summary. The intimate connection between emotions and respiration has significant implications in several research fields including psychology and HCI. Self-reporting of emotions, although practical, often comes with reliability and validity challenges due to social desirability bias, memory inaccuracies, and individual introspective abilities. A potential solution lies in complementing self-report measures with physiological data, such as breath analysis, providing a more objective perspective on emotional experiences. Voluntary control over respiration can be utilized to regulate emotions, as seen in practices like yoga and medical procedures for panic attacks. In the realm of affective computing, the analysis of breathing patterns could offer valuable biometric input for determining a user's emotional state, although discerning emotion-related factors from other influences remains a challenge. Non-invasive techniques, such as the use of radio ultra-wideband radar, show promise in achieving accurate emotion recognition based on respiration. Furthermore, understanding the relationship between breathing and emotions could significantly enhance the realism of virtual agents, by incorporating dynamic respiratory behaviors, which could even mimic users' patterns to exhibit empathy. However, the effectiveness of simulated respiration in enhancing the naturalness of virtual agents is still debatable, emphasizing the need for further research.

2.2. How emotions impact task switching and performance

2.2.1. Cognitive Load, Task Switching, and Human-Computer Interaction

2.2.1.1. Task Switching and Cognitive Load in Modern Life

Over the last two decades, individuals are increasingly confronted with various technological tools that permeate their daily lives, such as social media platforms, smartphones, and the internet. These advancements have facilitated a significant shift in human behavior, especially when it comes to multitasking and the management of interruptions. Multitasking is both prevalent and important in modern-day life and the various gadgets frequently interrupt task processing to cue individuals for a different task or goal (Oh et al., 2021). With smartphones and the internet, users often engage in task switching while on social media platforms, checking emails, or responding to messages, all while simultaneously attempting to complete other tasks. The interruptions people deal with can be broadly classified as internal and external interruptions, both serving as cues or drivers of switching away from the current task (Salvucci et al., 2009). Internal interruptions or self-interruptions refer to an internal decision to stop the current task and focus the attention on another one (Adler & Benbunan-Fich, 2013). This might sound familiar from daily life where the knowledge of the snack cabinet often tempts people to briefly stop working towards the current goal and have a snack. It is worth speculating what having a plethora of available distractions only a tab away from the screen can do to the temptation of self-interruption. External interruptions refer to events or stimuli that originate outside of the user's task or focus that can divert the user's attention away from the current task. These interruptions can come in various forms such as environmental cues, alerts, or notifications (Adler & Benbunan-Fich, 2013). The constant availability of distractors and a barrage of interrupting stimuli could have implications for the demand for cognitive resources, task performance, switch costs, and other cognitive mechanisms. It is crucial to explore the phenomenon of task switching in greater depth to better understand its implications on human-computer interaction. Adding to the literature on task switching can help researchers and designers devise more efficient interfaces and tools that support users in managing their digital lives. By understanding the cognitive demands of task switching and the impact of digital multitasking on users, it is possible to create technologies that cater to the evolving needs of modern users and help them find a balance between multitasking and focused work. Some of the more prominent theories on this process, as well as important concepts, will be explored in the following section followed by the implications and possibilities they can offer HCI.

2.2.1.2. Theories of Task Switching and Switch Costs

The process of transitioning from one mental task to another, task switching, has received much attention within psychological research. This interest can probably be attributed to the fact that people tend to have more challenges when shifting to a novel task as opposed to repeating the same one (Schmitz & Voss, 2014). Various theories on the underlying cognitive mechanisms of task switching have emerged over the years. Switch cost is a highly robust phenomenon that appears in task switching, both for reaction time (RT) and percentage of error(PE) (Kiesel et al., 2010). An example is the conventional task-switching paradigm has participants either rapidly shift between two tasks that are indicated by a task cue before the target stimulus or repeat the same task. The switching cost is then calculated by subtracting RT or the percentage of error PE of the repeat trial from the results from the switch trial (Hsieh & Lin, 2019). This is a commonly used method and it is unclear if results from this paradigm generalize to other protocols for investigating task switching. Although this method of measuring switch costs using RT and/or PE is prevalent (Hsieh & Lin, 2019; Liu et al., 2016; Rogers & Monsell, 1995) there is not a consensus in the literature as to why this switch cost occurs (Schmitz & Voss, 2014).

When an individual takes on a task they are said to adopt a mental task set that corresponds to the task. The task set refers to the cognitive functions and mental representations that allow an individual to act by the demands of a particular task. A task set must contain representations of both relevant stimuli and appropriate responses, along with their corresponding stimulus-response (S-R) associations. The S-R mappings differ in difficulty and are easy for highly learned tasks like reading a word but become more difficult for arbitrary tasks (Kiesel et al., 2010).

2.2.1.3. Task-set reconfiguration theory

The endogenous theory of task switching or task-set reconfiguration is a classic theory on task switching and the reason for the prevalence of the corresponding switch costs. The theory assumes that when switching to a new task the respective task set must be activated in memory before proceeding. This time-consuming process of reconfiguring the cognitive system is what causes the switching cost. This reconfiguration consists of goal-shifting, where the previous goal is replaced by the new one in working memory, as well as updating task-relevant information. For simplicity, it can be said that the task set is retrieved from long-term memory and loaded into working memory. The increase in PE and RT is due to failures in reconfiguring the task set on time (Kiesel et al., 2010; Schmitz & Voss, 2014).

2.2.1.4. Task-set inertia theory,

Another (exogenous) theory of task switching, the task-set inertia theory, proposes that the main reason for switch costs is the inhibition of the previous task set. After this inhibition is done it is possible to load the new S-R rules in working memory (Kiesel et al., 2010; Liu et al., 2016). An observation from task-switching research in favor of the task-set inertia theory is the asymmetrical switch cost for tasks with different levels of difficulty or dominance (Liu et al., 2016; Schmitz & Voss, 2014). When switching from a less dominant and easier task to a more difficult and dominant one the switching costs appear to be higher than if the switch is made in the other direction. To explain this asymmetry the theory argues that when individuals perform a more dominant task it is more difficult to retrieve the task set. This requires them to inhibit the more easy-to-establish task set, which individuals in turn have to overcome in the following easier task. This results in increased RT and PE in those trials. Another explanation simply states that the activation of the previous task set carries over to the next explaining the asymmetry (Liu et al., 2016). Finally, there is evidence from neuroimaging studies in support of the task-set inertia theory. A correlation between switch cost and an increase in task-irrelevant neural activation after the switch has been made. This supports the claim that something is lingering from the previous task that needs to be inhibited (Schmitz & Voss, 2014).

The two theories deal with different aspects of the process of task switching and both seem to have something to offer. It is hypothesized here that they might work in combination to a certain level where the switching cost is some combination of the time it takes to reconfigure the task set and realign the goals and the time it takes to inhibit what the previous task required. There is some evidence for this in the literature and the models have been combined into multiple-component models of task switching. The process is split up into phases, before and after the switch where both theories have a purpose (Schmitz & Voss, 2014). Before the switch is complete there is a task-set preparation where the reconfiguration takes place and after the switch, the S-R links are selected according to the constraints of the previous phase and the effect of the inertia (Schmitz & Voss, 2014).

*Summary. T*his subsection presents an overview of the endogenous (task-set reconfiguration) and exogenous (task-set inertia) theories of task switching, as well as the multiple-component models that combine the two theories. The discussion highlights that both theories may contribute to switch costs and task-switching processes.

2.2.2. *Cognitive Load Theory and Task Switching*

To outline the implication of the previous discussion on task switching and switch costs to HCI it is worth discussing the more general term cognitive load and cognitive load theory (CLT). Cognitive load refers to the mental effort required to process and retain information in short-term or working memory while performing a task. Information stored in long-term memory is in the form of mental automation schemas or knowledge structures that can facilitate the processing of known tasks, reducing the working memory required for the task (A et al., 2021; Kirschner et al., 2018). According to CLT, working memory is limited and only can deal with a certain amount of information at a time. Working memory capacity is required for all conscious cognitive behavior and overloading the memory, the cognitive load, reduces an individual's capability to process and perform a task (A et al., 2021). According to CLT, three types of cognitive load have an additive effect on working memory namely intrinsic, extraneous, and germane. Intrinsic load depends on the complexity of the task or material being learned, determined by the number of elements, their interactivity, and the proficiency of the individual. Intrinsic load is necessary for learning and problem-solving. Extraneous cognitive load results from the design or presentation of information elements that are not related to the task. In other words, they take up working memory with information that is not needed for the task. This is not advantageous for learning or problem-solving (A et al., 2021; Kirschner et al., 2018). The final type of cognitive load, germane load refers to the cognitive effort invested in constructing and automating mental. Germane load is influenced by the learner's prior knowledge and enables individuals to use learned strategies to make tasks easier to solve (A et al., 2021).

As previously discussed in this section task switching demands switching costs that occupy working memory. Switching costs are using the same limited cognitive resources as the three types of cognitive load making them closely related. If a task requires a high level of cognitive load, the individual's working memory capacity may be limited, leaving fewer cognitive resources available for task-switching processes (A et al., 2021; Schmitz & Voss, 2014). This can result in increased switch costs, as the individual may take longer to reconfigure the task set or to suppress irrelevant information for example a previous task set when transitioning between tasks. Here it is easy to connect the theories where the task-set reconfiguration prepares the individual and adds to the intrinsic load while the task-set inertia theory serves to reduce the extraneous load caused by the previous task. In the same sense switch costs contribute to cognitive load where the resources demanded for switching reduce the availability for task processing and solving.

*Summary. T*his subsection discusses cognitive load theory, the three types of cognitive load (intrinsic, extraneous, and germane), and their relationship with task switching and switch costs. It emphasizes the importance of considering cognitive load when designing HCI systems.

2.2.3. *Applying Task Switching and CLT Insights to HCI Design*

In the modern digital environment, there is a barrage of external interruptions, constant awareness of available distractions tempting people to self-interrupt, and an expectation to multitask throughout the day. There is no reason to believe that humans have evolved over the last two decades to increase the amount of working memory to manage this increase in cognitive load. That puts pressure on the technology to change to meet the needs and demands of people, to make sure that they can effectively interact with the technology within the constraints of their cognitive capacities. HCI can bridge the gap between the psychological understanding of cognitive limitations and the design of technology that aligns with these constraints, allowing a more seamless and efficient user experience. The importance of considering cognitive load when designing HCI systems has been made clear.

When designing HCI systems, it is crucial to consider cognitive load and the insights provided by CLT to create interfaces that facilitate efficient task switching while minimizing switch costs. As previously discussed there are tools available to help guide the development of systems that align more closely with human cognitive capabilities and enhance general user experience. In the following subsection, the tools offered by CLT and the task-switching literature will be briefly discussed in the context of HCI followed by examples of how they could prove useful.

To manage cognitive load the first step is important to manage the intrinsic load. Intrinsic load is based on individual proficiency in the task and increases with an increase of elements related to the task, their complexity, and how they interact with each other. When designing an interface this could be managed by breaking complex tasks down into simpler subtasks that include fewer elements and are more focused. This approach aims to manage intrinsic load in a way that sustains efficiency while preventing cognitive overload. Reducing extraneous load. Extraneous load, the task-irrelevant information that takes up working memory can be reduced by removing all unnecessary elements from the environment. Information and task interface should be clear, concise, well organized, and free from all distractions. Germane load should be supported allowing users to apply and acquire automatic mental schemas that help them deal with tasks more efficiently. To do this interfaces should build on the user's existing knowledge by offering adaptive paths that personalize the interface and cater to individual needs and skills. Finally, switch costs have to be minimized. Users must be supported in task-set reconfiguration and the suppression of irrelevant information. This could be done by providing clear and consistent interface elements and layouts across tasks and applications to help minimize the time and effort required for task-set reconfiguration Arranging tasks in a way that lowers switch costs by reducing the need for inhibition and creating subtasks that can be finished in order. Offering tools to manage notifications and alarms could similarly reduce switch costs. To sum up, this subsection outlines practical approaches to managing cognitive load and minimizing switch costs in HCI design. It suggests breaking complex tasks into simpler subtasks, reducing extraneous load, supporting the germane load, and providing tools to manage notifications and alarms.

The importance of considering switching costs and cognitive load when designing HCI systems has been discussed to help users manage their task switching. In the following sections, the influence of emotions on this whole area of task switching will be discussed in detail. Emotions affect task switching at every level and cannot be ignored when the theories of task switching and cognitive load are utilized in the development of interfaces and other systems.

Summary. The rapid integration of technology into daily life has increased multitasking and interruptions, making it crucial to understand the cognitive demands of task switching and cognitive load to improve HCI. The theories of task switching, cognitive load theory, and their practical applications in HCI design can help bridge the gap between psychological understanding of cognitive limitations and technology design, enabling more seamless and efficient user experiences. By applying these insights, designers can create interfaces that facilitate efficient task switching, manage cognitive load, and cater to the evolving needs of modern users.

2.2.4. Effect of Emotions on Successful Task Switching and Performance

The bidirectional relationship between emotions and physiology has been explored extensively. Equally important is the intricate interplay between emotions and cognition, which significantly impacts a wide range of cognitive functions. Just as emotions influence physiology, they affect cognition. Cognition is needed to produce emotional responses and emotion modulates and directs cognition to better interact with the world (Brosch et al., 2013). Both negative and positive emotions influence cognitive control and attention (Dreisbach & Goschke, 2004; Hart et al., 2010), memory encoding, consolidation, and retrieval (Levine & Pizarro, 2004; Phelps, 2004), perception (Brosch et al., 2013) as well as shaping rationality, value attribution and judgment (Pham, 2007).

The following section will further detail how the combined effects of emotions on various cognitive functions influence task performance and the process of switching between different tasks. There will be a brief discussion on how emotions affect each of the aforementioned cognitive functions with a simple base in neuroscience. The different effect of emotions of different valence and arousal on cognitive flexibility and performance is examined with a further look at how this can impact task switching.

2.2.4.1. Emotional Influence on Attention and Cognitive Flexibility

The connection between attention and emotion is a widely studied topic. People are surrounded by an immense amount of sensory information, much more than what can be processed by limited brain resources. It is, therefore, necessary to filter out and prioritize what sensory information is further processed, to pay attention to certain elements at the cost of others (Brosch et al., 2013; Tyng et al., 2017). Emotions have been shown to significantly alter the way attention filters and prioritizes sensory information. Some emotions impact the scope of attention. Negative emotions tend to narrow the focus of attention and prioritize a single mode of action while positive emotions widen the scope and encourage a more open mode of action like explore or play (Dreisbach & Goschke, 2004; Jeon, 2017). How the scope of attention has implications for task switching will be discussed later in the section.

Various behavioral tasks have demonstrated that emotions impact the selection and facilitation of the information that people pay attention to. Compared to neutral stimuli, emotional stimuli draw attention quickly and hold attention for a longer time. This has been demonstrated with both visual and auditory search tasks where an emotional target is detected faster among distractors than a neutral one (Brosch et al., 2013). This sort of emotional cueing impacts the perception as well as the attention with studies suggesting that emotions change the content of the perception across multiple modalities creating a kind of attentional bias (Brosch et al., 2013). To give an example this effect could present itself when a parent picks up their child at a day-care center crowded with children and can hear their voice through the noise. On the flip side of this function some studies have shown that when competing against each other the presence of emotional stimuli will limit cognitive control by disrupting the ability to pay attention to neutral task-relevant stimuli (Hart et al., 2010). FMRI brain imaging studies have given support for this view of the emotional impact on attention. More robust neural activation is seen in early sensual areas when confronted with emotional stimuli when compared to neutral. A stronger signal suggests that an emotional signal would get preferential access to higher brain areas for further processing and awareness. Interestingly further FMRI studies have suggested that neural circuits around the amygdala are, at least partially, responsible for this prioritization of emotional stimuli before it reaches conscious attention and irrespective of where the attention is focused (Brosch et al., 2013; Phelps, 2004; Vuilleumier, 2005). A study that compared healthy subjects with patients with bilateral amygdala damage showed that the healthy subjects showed more attention and perception to emotional words than the other group (Tyng et al., 2017). This could explain both why it can be difficult to convert attention away from emotional stimuli but also why people are quick to identify it.

2.2.4.2. Cognitive Flexibility and Task Switching

How emotion affects attention is an important factor when trying to understand how the affective state both shapes and is shaped by the ability to switch between tasks. Cognitive flexibility is a crucial aspect of attention and refers to the ability of attention to adapt to the changing environment(Cools, 2015). In other words, the ability to switch thought from one concept to another and to flexibly update cognitive and behavioral strategies accordingly (Cools, 2015; Magnusson & Brim, 2014). The switching of strategies entails inhibiting the previous task to manage the performance demands of the next. An example of a measurement of the level of cognitive flexibility is the Wisconsin card sorting task. The task involves sorting cards based on continuously changing rules they have to deduct based on feedback. If participants preserve a rule for long is taken as a marker of an inability to inhibit it and replace it with a new one. People with high cognitive flexibility are more successful at inhibiting the previous rule and switching to a new one (Liu et al., 2016).
As previously mentioned emotion has a significant impact on the scope of attention. Positive emotions have a broadening effect on the scope and thought-action repertoires while negative ones tend to narrow both (Fredrickson & Branigan, 2005). This suggests that positive emotion would facilitate smoother task-switching and vice versa. The effect of emotion on cognitive flexibility has been widely investigated and, in line with studies on the scope of attention, both negative and positive emotions have been found to have a significant effect on the ability to switch between tasks.

2.2.4.2.1. The Broaden-and-Build Theory of Positive Emotion

The broaden-and-build theory of positive emotion is a psychological framework that has been used to predict the effect of a subset of positive emotion on multiple cognitive factors. Positive emotions like joy, interest, and love broaden cognitive and behavioral repertoires and encourage behaviors such as exploration, play, and savoring. In other words, they increase the range of possible thoughts and actions that spring to mind in a given situation. A less important aspect of the theory for the current research is the building element. This state of open-mindedness accompanied by positive emotion supposedly helps build intelligence and knowledge and other personal resources (Fredrickson, 2004). In line with the broaden-and-build theory, positive emotions are expected to improve cognitive flexibility and task-switching abilities, as individuals become more open and receptive to engaging with new tasks. This effect was shown in Dreisbach and Gosche's 2004 study where positive affect induced by images had a significant effect on the ability to disengage from a task-relevant stimulus category and switch to a new one. They also found an increase in distraction following positive emotion (Dreisbach & Goschke, 2004). The impact of emotional valence on task switching has been a subject of debate since the 2004 study, and the influence of the broaden-and-build theory on cognitive flexibility remains contested. This will be discussed in more detail in the following subsection.

2.2.4.2.2. Negative Affect and Cognitive Flexibility

There is evidence for negative affect having detrimental effects on cognitive flexibility and task switching (Fredrickson & Branigan, 2005; Hart et al., 2010; Paulitzki et al., 2008). Negative mood has been shown to hinder relaxation of cognitive control so more time is needed to adjust after switching tasks (Hsieh & Lin, 2019). A study on task switching speed where participants switched between a neutral task involving numbers and a task determining the texture of a spider. Participants that scored higher on a questionnaire estimating their fear of spiders were both quicker to switch to the task involving the spider and slower to switch from it to the digit task. These results are in line with previous research on the interaction between task switching and aversive stimuli (Paulitzki et al., 2008). These findings fit the previous discussion on cognitive flexibility where it would be expected that inhibition of the strategies of a fearful task is slower than that of a neutral one. A theory often used to explain this effect of negative emotions on cognitive flexibility is that negative emotions like fear, anger, and anxiety demand highly specific action tendencies. These are urges for a certain behavior combined with clear physiological responses optimized for that behavior (Fredrickson & Branigan, 2005). This fixated focus on that single action, while omitting distractions and other options, such as the next task. This explanation of the interplay between negative emotions and cognitive flexibility is tempting as it neatly fits an evolutionary view on emotions where anger can be explained by fight and fear by flight. It is still much up for debate if and how negative emotions affect task switching and that is especially true for the valence dimension. An arousal hypothesis predicts that, irrespective of valence, stimuli of high arousal strengthen the components of the current task set, causing it to be less prone to interruption from new response conflicts and more stable. This in turn decreases cognitive flexibility and slows task switching (Demanet et al., 2011).

2.2.4.2.3. The Role of Emotional Valence and Arousal in Task Switching

High arousal positive emotions are often thought of as high in approach motivation, desire, for example, would cause similar detriment to cognitive flexibility as fear and anger. The positive emotions on the lower end of arousal and approach motivation like happiness would on the other hand enhance flexibility (Wang et al., 2017). This emphasizes the need to control for the arousal dimension of emotion if the effect of valence on cognitive flexibility is to be studied. A 2011 study by Demanet et al. had participants perform a simple voluntary task-switching paradigm after inducing an affective state using images. Controlling for arousal they did not find any effect of emotional valence on task switching or switch cost. They did find that a positive affect increased global performance and suggested that was due to an enhanced ability to multitask. Affective arousal did impact switch cost with increased arousal slowing task switching. The results of this study are at odds with previous studies on affective valence on task switching. The researchers suggest that this is due to a difference in the tasks and their control of the arousal dimension. They also note that they mixed positive and negative affect in the same block, which they think could cancel out the effect of valence (Demanet et al., 2011).

A 2019 study by Hsieh and Lin had participants watch an emotional film before performing task switching. They measured cognitive flexibility in two ways, on a long and short-term scale. Switching cost, the difference in reaction time and errors in a task switch trial compared to a task repeat trial and fade-out effect, the time it takes for task performance to increase after a task switch. They found negative moods, controlling for arousal, reduced switching costs when compared to positive or neutral moods. This was particularly true in incongruent trials. Negative moods increased the fade-out time, meaning participants took a long time to adjust to the task after switching(Hsieh & Lin, 2019). The impact of negative mood on fade-out time is on par with previously discussed theories, where it is more difficult to inhibit the previous task when negative affect is prevalent.

A third study from 2017 had participants look at images of different emotional valence and perform a task-switching paradigm in an fMRI machine to investigate underlying neural mechanisms. Due to the nature of the study it suffered from a lack of participants (n=19). They found that a positive emotional state lowered switch cost and reduced brain activity in relation to task switching while the opposite was true for negative emotion. They suggest that negative emotions require more cognitive resources thus requiring more attentional control to accomplish a task. The emotional stimuli used in the study were mild-arousal so the authors conclude that the effect is due to emotional valence(Wang et al., 2017).

The three studies mentioned above suggest that the impact of negative emotion on task switching is far from conclusive and there is much need for further research. It is unclear if valence has an effect altogether and if it does there is evidence for negative emotions to both enhance and suppress task switching and cognitive flexibility. The studies did vary slightly in their choice of tasks and how they were presented. Although they all account for arousal they use different levels of arousal which opens up a possibility for an interaction between the dimensions where positive emotions could be differently affected by arousal than negative and that an effect of valence is only present at certain strength of arousal.

Summary. The connection between attention and emotion plays a significant role in cognitive flexibility and task switching. Emotions alter the way attention filters and prioritize sensory information, with positive emotions broadening the scope of attention and negative emotions narrowing it. Cognitive flexibility, crucial for adapting to a changing environment, is impacted by both positive and negative emotions. The broaden-and-build theory suggests that positive emotions encourage open-mindedness and improved task-switching abilities. However, the relationship between emotional valence and cognitive flexibility is still contested. Negative affect has been found to hinder cognitive flexibility and task switching, with various theories explaining this effect. The influence of emotional arousal is also an important factor to consider when studying the impact of emotions on task switching. Further research is needed to clarify the effects of emotional valence and arousal on cognitive flexibility and task switching.

2.2.4.3. Task Switching in the Domain of HCI

The association between emotion and task switching has been extensively investigated in traditional psychological research, as outlined earlier. This crucial relationship is of significant interest within HCI, where innovative methods for exploring emotions and task switching are being developed. As users interact with increasingly sophisticated interfaces and multitask in digital environments, it becomes essential to comprehend the complex dynamics of emotion and attention when shifting between tasks. It is of interest to study the interplay between users' emotional states and their task-switching behavior, which could inform the development of more intuitive and responsive systems.

An innovative study from 2021 investigated how perceived cognitive and emotional in the context of flow impacted performance and task-switching behavior (David Bowman et al., 2021). Unlike the aforementioned psychology studies, this study did not use images or video to induce affect but altered game mechanics. 123 participants played an asteroid-like video game where they blew up asteroids and collected crystals. The researchers were able to manipulate the difficulty and rewards ratio of the game creating three scenarios: low effort/low reward that was supposed to induce boredom, high effort/low reward that induced frustration, and finally a balance of the two that was supposed to have participants in a flow (David Bowman et al., 2021). Flow is a state frequently used (but not limited to) to describe an optimal focus to maximize performance in an activity. This state is usually described as pleasurable and enjoyable (Chen, 2007). To measure these states the researchers had both behavioral data logged from the video game and self-report of enjoyment.

An interesting twist on the task was that every once in a while there was an opportunity to switch to another task, small sprites appeared on the edges of the screen prompting the participants to do a keystroke on a keyboard. The reaction times to these secondary tasks were used as a measure of cognitive resources allocated to the primary task with shorter reaction times indicating lower allocation of resources to the primary task. Unsurprisingly the bored, low-effort group had the lowest reaction times. The task was simply not demanding so they could multitask with ease. The flow group was moderate in their response times suggesting that the bulk of their cognitive resources were focused on the primary task. Interestingly the group that had the greatest improvement in reaction times was the frustrated group. This was interpreted as them getting instant gratification from completing the secondary task causing them to fully task switch instead of trying to multitask. This switch is referred to as cognitive offloading where users shift engagement from a frustrating task to another one to cope with the cognitive load (David Bowman et al., 2021). A problem with this study is that although they measured enjoyment and the frustrated group had the lowers score there (David Bowman et al., 2021), they primarily focused on how demanding the task was. It would be interesting to see how the manipulation of game mechanics could be used to induce emotional states to see their effects on cognitive load and task switching.

There is more evidence that an imbalance in the demands of a task and the ability of the individual increases the likelihood of task switching. When a task is too demanding it induces negative emotions like frustration or exhaustion and if it is too easy it can trigger boredom or stimulation (David Bowman et al., 2021; Oh et al., 2021). These emotions related to this imbalance are likely to trigger self-interruptions, although that is especially true for negative emotions. This tendency to self-interrupt during a frustrating difficult task is likely due to the high cognitive resource demand and coping mechanisms to retain those resources (Oh et al., 2021).

A 2019 study on task switching in a smartphone environment collected screenshots from the smartphones of participants every five seconds. They ended up with over two million screenshots that they used to train a model to predict task switching using six features; visual stimulation, cognitive load, velocity and accumulation, sentiment, and time-related features. The researchers wound up being able to predict the task-switching behavior of participants with 77% accuracy. The most predictive factors were visual stimulation, cognitive load, and the velocity and accumulation of those features. Most importantly for the current research sentiment was deemed unimportant in the prediction of task switching with positive and negative emotion having an "importance" score of 3.3 and 3.6 respectively compared to 34.8 with the highest rated feature. They conclude that this is due to the fast-paced nature of task-switching behavior in an increasingly visual user experience that leads to visual features weighing more than text-based, content features (Yang et al., 2019). These results suggest emotional valence is a poor predictor of task-switching behavior in a smartphone environment. In the context of the study previously outlined it would suggest that the task-switching behaviour had more to do with cognitive load than the emotion of frustration. These findings are unexpected as when competing for cognitive resources, emotional information has been found to take priority and disrupt the ability to focus on a non-emotional cognitive task (Hart et al., 2010;

Vuilleumier, 2005). Emotional stimuli have also been shown to draw attention for a longer period when compared to non-emotional content (Brosch et al., 2013). It is, however, worth mentioning the uniqueness of the smartphone user experience. It is incredibly visual and where enticing visuals are used to compete for the user's attention. It might be that the emotional content of the tasks at hand is simply overshadowed by these stimuli. Another explanation might be that in the context of a smartphone environment users have learned to prioritize their attention more to visual elements or novelty of tasks instead of their emotional resonance. Finally, to assess sentimentality they analyzed text from the screenshots (Yang et al., 2019). There is no guarantee that the content on the screen evoked any emotion in the participants and possible that the visual stimuli had a more profound effect on emotion.

Summary. The association between emotion and task switching has been a topic of significant interest in both traditional psychological research and HCI, with innovative methods emerging for exploring this relationship. Studies of in-game mechanics and smartphone environments reveal the complexity of the interplay between users' emotional states and their task-switching behavior. Findings suggest that the balance between task demands and individual abilities, as well as the nature of the environment, can influence the impact of emotions on task switching. Understanding this relationship is essential for the development of more intuitive and responsive systems, particularly as users interact with increasingly sophisticated interfaces and multitask in digital environments.

2.2.4.4. Emotion, Memory, and Task Switching

2.2.4.4.1. The Effects of Emotion on Memory Processing

The connection between memory and emotion has been the topic of much research over the last decades, especially on whether emotions enhance or skew memory and if memory recall differs depending on the current emotional state (Levine & Pizarro, 2004). Much of the research has been focused on the effect of emotion on long-term memory. That will be discussed briefly but is for the most part outside the scope of this review. In simple terms memory processing can be described as a three-stage process; encoding, consolidating, and recalling. The stages respectively process sensory information when it is perceived, store the information in the brain, and finally remember the stored information at a later time. Emotion can influence all three stages at some point (Brosch et al., 2013). As previously mentioned emotion can influence attention and perception by emphasizing emotional stimuli. As higher priority is given to the emotional stimuli memory encoding of that stimuli is enhanced (Phelps, 2004; Tyng et al., 2017). The consolidation process is relatively long and newly formed memories are fragile, and during this time it is suggested that memories can either be weakened or strengthened and that emotion modulates this process (Brosch et al., 2013; Phelps, 2004). Finally, and most importantly for the current research, emotion affects memory recall in a complex manner. It is common for people to describe highly emotional memories, a car crash for example, as vivid and detailed. Interestingly, when people are tested on this detail these memories have the same deterioration as neutral memories but people simply are more confident they are correct. It has been theorized that this confidence in critical memories is adaptive in an evolutionary sense. In a dangerous time-critical situation people search their memories for a similar event to guide their actions and a doubt or hesitation to use that memory can be costly to the situation (Brosch et al., 2013). Although emotional memories do not seem to be as vivid as many think other studies show that emotional intensity predicts greater memory vividness (Levine & Pizarro, 2004).

2.2.4.4.2. Emotional State and Memory Recall

Recall is highly dependent on the emotional state (affective state) at the moment of recollection (Eich et al., 2008). Laboratory studies on humans demonstrated this in a "brute force" manner by showing that injecting stress hormones (epinephrin) into the amygdala at the time of consolidation increases recollection of emotional memories (Levine & Pizarro, 2004; McGaugh, 2000). The fact that emotion and memory are intertwined is highly relevant for understanding how emotions can impact task switching. There are various ways that different affective states influence how people access memories. Emotional states affect what information is recalled both from long-term memory and recently formed memories (Levine & Burgess, 1997). Unsurprisingly, the difference in memories accessed depending on emotional state commonly follows what information is relevant to the purpose of each emotional state. Happiness or joy is associated with openness to novelty and paying attention to a wide range of information. Happiness suggests that there is no immediate need for problem-solving and one is free to explore new goals or tasks (Levine & Burgess, 1997). Negative emotions such as anger or fear are more focused, they have a defined purpose or an urge to action, to remove an obstacle or avoid a threat (Fredrickson & Branigan, 2005). Negative affect with high arousal (anger) has been shown to enhance recollection of specified aspects of events. This is seen in eye-witness testimonies for example. Conversely, positive emotions, feelings of happiness, or joy seem to enhance the recollection of peripheral information (Talarico et al., 2009). A study of 263 undergraduates randomly assigned them a grade from A to D on a quiz. Right after the quiz they heard and recalled a story. After describing their emotional state they were asked to recall the narrative of the story. The participants that described themselves as happy were more likely to recall more of the story as a whole. The self-described angry students remembered less of the story but had enhanced recollection of goals in the narrative and sad students focused on outcomes. This study shows that the effect of an affective state on recollection impacts newly acquired information as well as long-term memory (Levine & Burgess, 1997).

2.2.4.4.3. Affective Priming and Information Processing Strategies

Affective priming is the tendency to recall mood-congruent information. When negative or positive emotions are induced in people, they tend to selectively recollect events that match their respective affective states. This has been demonstrated in participants that kept a diary for weeks and were then asked to recall details from it. Affective priming has also been shown to impact people's recollection of their social behaviors and those of others (Eich et al., 2008). Even more relative to task switching, the affective state has been associated with different information processing strategies. Consistent with the description of happiness above, people tend to rely more on general knowledge and low-effort heuristics for information processing. Studies suggest that individuals who are experiencing feelings of happiness tend to rely more on their general knowledge, stereotypes, or other simplified decision-making strategies when evaluating arguments or making social judgments. Conversely, individuals experiencing negative emotions rely less on general knowledge but instead evaluate information carefully and systematically using more costly information processing strategies(Levine & Pizarro, 2004).

Emotional memory has important implications for task switching. Emotions induced by the first task may have much impact on how the subsequent task is approached, what memories are accessed to assist with the task, how it is perceived, and what strategies are used. In the other direction, emotions evoked by the second task may influence how the former task and the evoked emotion are remembered.

Summary. The connection between emotion, memory, and task switching is a complex and intertwined relationship, with emotions affecting various stages of memory processing, including encoding, consolidation, and recall. Emotional state at the time of recollection plays a significant role in memory recall, with different emotions influencing the type of information that is accessed. Affective priming demonstrates that people tend to recall mood-congruent information and emotional states can be associated with different information processing strategies. Understanding this interplay is crucial for comprehending how emotions induced by tasks can impact task switching, influencing the approach, perception, memory access, and strategies used for the subsequent task. Additionally, emotions evoked by the second task may affect how the former task and its associated emotions are remembered.

2.3. Emotion-Based Cognition in HCI and Task Switching

2.3.1. Emotionality and Landscape Preference Model (LPM) in HCI

The previous subsections have detailed how emotions have a significant impact on almost every aspect of task performance and switching. They impact attention, most stages of memory, cognitive flexibility, attitude toward novel information or stimulus, and even the processing strategies likely to be deployed. This broad effect of emotions on cognitive processes and attitudes underlines the importance of taking them into account when designing HCI. It is valuable to know how the emotions evoked by these systems impact how users interpret, react, or use the next system or interface. As an example, a poorly designed, confusing interface could evoke negative emotions of frustration or agitation. The effect of these emotions can have various implications on how the user interacts with the following system, whether it is finishing a purchase in an e-commerce store or playing an adventure game. Another interesting aspect is that evoking certain emotions with an HCI system might be appropriate for a particular follow-up task while other emotions would be a better fit for a different task. There might be a difference between how a system would evoke emotions before users interact with an online bank and a first-person shooter video game. This section explores how the knowledge of emotions in relation to cognition can affect the appraisal and the emotionality of a following task in the context of HCI. Emotionality in relation to interfaces will be explored through the lens of the landscape preference model LPM) and how it can be used to make predictions on how emotions could prime users for the subsequent task (K. Lee et al., 2019).

2.3.2. The Four Elements of LPM

Individuals tend to use their immediate emotional responses to shape their preferences for the environment. Emotions can be thought of as an introspective source of information (K. Lee et al., 2019). Kaplan described four elements that primarily impact human preference in a given environment: Coherence, Legibility, Complexity, and Mystery (Kaplan, 1975; K. Lee et al., 2019). These four concepts lie in a two-dimensional matrix with one dimension representing the essential goals of sensemaking and exploration. Sensemaking is the drive to understand what is going on at a given moment and exploration corresponds to the interests sparked by the anticipation of new experiences or knowledge. The other dimension further splits these goals into immediate and inferred elements depending on how much cognitive processing is required to assess the information (K. Lee et al., 2019).

The first goal, sensemaking, consists of two elements: coherence and legibility. Coherence describes qualities that make a scene easy to organize and comprehend at an immediate level. In contrast, legibility refers to the conditions that hint at making sense of a scene at an inferred level. For example, markers that help navigation or signs that contribute to a sense of safety and control. The other essential goal, exploration, is satisfied through the perception of complexity and mystery. Complexity describes the qualities that bring richness to a scene at an immediate level, such as the amount of activity or visual engagement. Mystery represents the promise of additional information that could be uncovered by further exploring the scene (K. Lee et al., 2019).

2.3.3. Regulatory Focus and Task Transitions in HCI

Although this framework appears abstract it has been successfully adapted to HCI research including website design and measurement tools for online user behavior (K. Lee et al., 2019). The LPM framework has been used to create a scale that accurately predicted impressions of websites and how likely users were to return. Similar uses have been found in online shopping environments where it can explain user’s perception of the shopping environment and explain a large portion of the variance in purchase intention (Demangeot & Broderick, 2007; K. Lee et al., 2019; Y. Lee & Kozar, 2009). There is a reason to believe this model can be helpful for the current research as it can be easily translated into the valence-arousal model of emotion and allows describing the affective states of individuals in terms of goal-driven behavior. Through that lens, the primary function of emotions is to give introspective feedback that signals success or failure in pursuit of personal goals. To understand emotions it is important to identify the motivation that triggers them (K. Lee et al.,

2019). To frame this motivation K. Lee et al. point to Regulatory Focus Theory which states that humans have two fundamental needs, security, and nurturance that they fulfill by self-regulation, either through prevention focus or promotion focus. The focus state is influenced by the situation, or task, at hand. In a promotion-focused state, individuals display eagerness and work towards ideal goals aimed at gaining something. Conversely, in a prevention-focused state
individuals are vigilant, framing goals towards avoiding loss, so-called ought goals (K. Lee et al., 2019). The emotions accompanying those goals map onto the circumplex model of affect, the valence-arousal model where achieving an ideal goal evokes high arousal-positive valence emotions like joy and excitement while failing those goals produces low arousal-negative valence, sadness, or gloom. On the other hand, achieving an ought goal users feel low arousal-positive emotions such as relaxation or relief and high arousal-negative valence if they fail to achieve them, agitation or anxiety (K. Lee et al., 2019). These map onto the valence arousal as continuums of promotion and prevention emotions respectively as can be seen in Figure 1.1 (K. Lee et al., 2019). For HCI design sensemaking, decreasing complexity, increasing legibility, and comprehending the context can be thought of as a prevention goal. Exploration on the other hand can be considered a promotion goal where the purpose is to reach out for more excitement and seek more positive stimulation. Design choices that influence sensemaking should evoke emotions on the restless-calm continuum and those that incite exploration on the joy-sadness continuum (K. Lee et al., 2019).

2.3.4. Emotion-Based Task Switching

Understanding the emotional states and the regulatory focus of users is helpful for HCI research. It has implications for task switching where it underlines the importance of optimizing the order of tasks and how certain tasks could be made to follow an interface or an activity that builds up the appropriate emotional state to match its demands. When a user is interacting with a website to pay taxes or access email it can be thought of as an ought coal. The road to the goal should be optimized for sensemaking, free of distractions, and offer a direct route to prevent negative attitudes. With a system that revolves around enjoyment, a gaming interface, or a streaming service users should be more sensitive to design that encourages exploration and drives positive attitudes. This can mean a multitude of distractions ready to explore and an induced joy and excitement for the entertainment to come (K. Lee et al., 2019). This extension of the landscape preference model adds to the list of possible effects emotions can have on task switching and the interpretation, performance, and emotional response of the subsequent task. A smart system that detects emotions, and adapts accordingly could use this information to adjust interphases dynamically depending on prevention or promotion-related emotions.

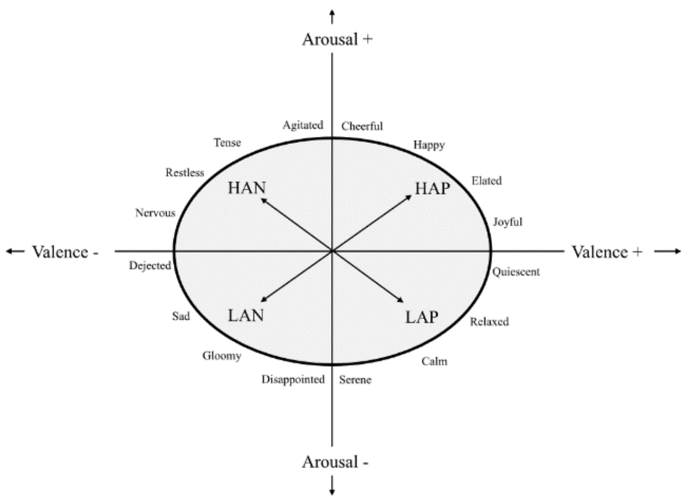


Figure 1.1. The circumplex model of affect with the added promotion and prevention focused continuums (K. Lee et al., 2019)

Summary. Emotions significantly impact task performance and switching in Human-Computer Interaction (HCI), influencing users' cognitive processes and attitudes. The Landscape Preference Model (LPM), encompassing four key elements: Coherence, Legibility, Complexity, and Mystery, provides valuable insights into these interactions. Aligning LPM with the valence-arousal model of emotion helps to understand users' affective states in goal-driven behaviors. Regulatory Focus Theory further elaborates this understanding by differentiating between security-driven (prevention) and nurturance-driven (promotion) needs. These insights can optimize task sequencing in HCI design, catering to users' emotional states and promoting beneficial attitudes and performance. The LPM model thus significantly contributes to a comprehensive understanding of emotions' role in HCI.

3. Methodology

The following section provides a thorough overview of the research design, procedures, and tools used in the experiment. The primary objective of the study was to explore how users' emotional responses and breathing patterns are influenced while interacting with a specifically designed online clothing store. This was achieved by having participants navigate the website while their mouse clicks, breathing patterns, and screen interactions were recorded and analyzed. The research setup involved a controlled environment where participants performed tasks on a laptop equipped with various tracking tools, including AutoHotkey scripts for mouse logging, OpenSignals for breathing measurement, and OBS Studio for video recording. Participants were required to perform specific tasks on the website, such as character creation, browsing the store, and account creation, with the latter being intentionally designed to induce (negative) emotions through a challenging interface, based on qualitative feedback from pilot studies. This section details each step of the study's procedure, from the participant setup, including wearing a breathing belt and engaging in a calming breathing exercise used for calibration, to navigating the website and completing various tasks, one of them being a list of self-assessment questions about their emotional status at varying points. The comprehensive approach to data collection, including a video recording of the screen activity, breathing data measurement, mouse click logging, and survey responses allow for a detailed analysis of the participants' interactions with the interface and their consequent emotional and physiological responses.

3.1. The experiment

3.1.1. Participants

The study was conducted with a total of 19 participants, comprising a group of 8 females and 11 male participants. Participants were chosen based on convenience and through advertising in a student housing building and on a WhatsApp group consisting of students of Human-computer Interaction. In the advertisement, students were directed to a survey where they could pick a time slot that suited them. Participants were in the age bracket of 25 – 35. The dataset from one participant was omitted due to a problem with the gathering making it impossible to normalize it in the same way the other sets were normalized. This sample size, while modest, was considered sufficient to achieve the objectives of the study, particularly given the exploratory nature of the research and the depth of data collected from each participant, with the hope that a similar study can be conducted with a greater sample size in the future.

3.1.2. Procedure

The study takes place in a quiet room or an area, this could vary from participant to participant. Some took place in offices or rooms in the Science park, while others took place in participants' homes, at the home of the researcher, or a library. Regardless of location care was taken that the area was quiet and without distractions. A laptop computer with an external mouse is located on a table with a chair in front. The computer has been prepared beforehand by opening the two mouse logging AHK scrips that track mouse clicks and timestamp them. Opensignals has been opened and connected to the breathing belt via Bluetooth and OSB studio is open and running in the background with the preset made for the study. Microsoft Edge is open and all cookies and history have been cleared, all tracking and "auto-filling of forms" has been disabled in the settings. The website for the study is open on the screen with the consent form ready.

When the participant sat down in the chair they were introduced to the breathing belt, how to put it on, and how to adjust its size. The researchers showed them how to put it on and told them that it should be fairly tight around the breast area right under the so-called nipple line. Care was taken that there is no twist on the belt and that is it comfortable. After they put on the belt they were pointed to the computer. They were told that they will go through an online store and the procedure will be explained in detail during the experiment. When the researcher had turned on the recording of the breathing signal the participant was asked to do a “box breathing exercise” that consists of inhaling deeply and counting to three, then exhaling deeply and counting again, they should repeat this 5 times. This exercise has the purpose of regulating the breathing of the participant and calibrating the breathing signal. The video recording of the screen is started and the website was set to fill out the screen of the laptop. Participants were now asked just to breathe normally while they read the consent form (appendix 1) and to continue when they are ready. They were asked to refrain from talking and use the mouse to traverse the interface and the keyboard for writing. The experiment with instructions took around 15-17 minutes.

3.1.3. Study Setup Step-by-Step

1. Switch keyboard to English
2. Open AHK script mouselog
3. Open AHK script mouselog on screen
4. Open Opensignals and start recording
5. Open OBS studio and start recording
6. Open Microsoft Edge and clear cookies for the survey to start at the right stage
7. Open the website and enter full screen
8. Greet the participant and explain the procedure and the breathing belt
    a. This belt will track your breathing while the experiment is running. It should go right below your nipple line. I’ll show them on myself and explain that this might feel a little weird (based on comments from the pilot studies). You should sit and try not to focus on anything else than the screen. You will get your instructions with the tasks of the experiment in the introduction following the consent form, please read them well. If you have any questions now is a good time. Please only use the keyboard for writing and the mouse to click buttons and traverse all forms.
9. Allow them to put the belt on themselves
10. ask them to box breathe to calibrate the signal
11. Ask them to only use the mouse to traverse the website
12. ask them to read the terms while breathing normally and tell them I will leave the proximity.

3.2. Tools and methods

3.2.1. Equipment

The experiment is run on an ACER Nitro 5 laptop computer running Windows 11. There is an external Razer Naga Trinity mouse that participants are asked to use. The experiment begins with the website being displayed on full-screen mode using Microsoft Edge as a browser where all cookies and saved data have been cleared and form suggestions and auto-fill have been disabled.

3.2.2. Website

3.2.2.1. General about the website and its purpose for the experiment

The website aimed to look and feel like a typical and modern website. It should be a familiar experience to browse it and make the users feel like they are doing something they are safe with and that should not provoke any negative emotion. It was supposed to give the users the impression that they were in charge and making their own way through it. It was decided that the website should appear to be an online clothing store with various items for the user to choose from. Every user would have an identical experience with the first two pages of the website, the first informing them of their rights and to get consent for the data gathering, and the second would be the beginning of the first task. It should lead all users to the same route with identical problems with the hope of provoking some emotion in the participants. After the experimental task, users should be redirected to a Qualtrix survey to answer questions about their experience, serving as the second neutral task of the experiment. The body of the website (the shop) allowed users to explore the site just as if it were a real store with a location and a mailing list. Because of the task they were given (to pick out an item of clothing and buy it) every page had a link to the store or some clothes available to click bringing them to the next account creation page that forced participants down an identical route. Appendix 2. has images for the important parts of the website.

3.2.2.2. Discussion about Duda

To achieve the familiar and typical feel of the website it was decided to use one of the most popular website builders available. Duda is a web design platform that powers over a million websites on the internet and offers simple and easy-to-use layouts for various types of websites https://www.duda.co/. The layout for a clothing store was chosen for its simplicity and pleasant look and feel. This would serve as the body of the website after various modifications. The layout was left functionally intact, but some images, texts, and links were changed to serve the purpose of the experiment. Below, each page of the website will be briefly explained with a focus on its purpose for the experiment.

3.2.2.3. Quick Overview of each Page

Each page will be discussed in a way deemed sufficient for the experiment. This means that some of the pages leading from the body of the store page will be left out since they only exist to give the website a more realistic look and serve no functional purpose other than to lead the participant to an item to add to their cart. The website in full can be seen at https://preview.duda.co/preview/4acb20e0 (*Duda | The Professional Website Builder You Can Call Your Own*, n.d.).

3.2.2.4. Consent Page

The consent page is what will greet the participants at the beginning of the procedure. It consists of a consent form (see Appendix 1) and a check box that logs their consent in a CSV file with a timestamp. As the user clicks the continue button task 1 (character creation) starts.

3.2.2.5. Information and Character Creation

This page is intended to be emotionally neutral for the participant. They read some text explaining the experiment in more detail and set them up with a task on the shopping website to come, the instructions can be seen in Appedix 1. They should create a persona using the form below where they are to give that persona a name, contact information, and some personal background like career or hobbies. They should then approach the following shopping task with that persona in mind. This is hypothesized to be a relatively neutral task and serves to create disposable user information to be used in the following task to avoid gathering any personal information and to try to mildly distract the participants from the main purpose of the experiment. The instructions given ask users to browse the web store looking for an item of clothing their persona would like to purchase and buy that item. They are asked to remember the details about their persona because they will have to use them at a later time in the experiment. No written information from this task is stored. Having filled out the form they can proceed to the store page.

3.2.2.6. Website Body

The website body consists of five pages: home, shop, about, lookbook, and shipping & returns. These pages were available as a template from Duda for a non-existing web store that sells clothes. The pages “about” and “shipping & returns” did not have a direct purpose for the experiment but served to make the website look more natural and the experience more realistic in case participants would explore the website. All text, notes, locations, and contact information on the site are made up. The “home” page is the first thing users see after they leave the character creation task. The purpose of that page is to have the store appear legitimate and lead users to select an item to purchase which brings them to the next part of the experiment. The first visible thing is a button leading to the “shop” page, there is a short text about the store followed by image links to the “lookbook” both of which lead users towards available clothes that bring them to the next part. After the user has clicked on an image of an item they have the option to "put in cart". When clicked a pop-up appears telling users to create an account to purchase that item. This brings them to the first "account creation" page.

3.2.2.7. Account Creation 1 – The Loops and Popups

Users are asked to create an account by filling in a form using their persona's details. This includes name, email address, password, repeated password, and date of birth. The form also has two drop-down menus selecting gender and for whom the user is planning to shop. The participant has to check a box to agree to share their responses and click "submit". None of the information is stored anywhere and is unreachable after the experiment. When the "submit" button Is clicked a popup appears asking users to prove they are not a robot by solving a Captcha that consists of skewed and extremely difficult-to-read letters that the user should write down below and click "submit" again. No matter what the user writes another popup will appear asking them to solve another Captcha, equally difficult and this time followed by a notification in red letters that "you have 2 more attempts". No matter what is written a third popup appears with a third difficult Captcha and a "you have 1 more attempt" notification. Just like the previous popups it does not matter what is written but this time it brings up a fourth popup saying there is an error with the information written in the form "Some information is missing or your password does not include one capitalized letter, one symbol, and at least 8 characters". The only way to continue is to click "return" which brings users to the second account creation page effectively refreshing the page and removing all information entered in the form.

3.2.2.8. Account creation 2 – Looks the Same but Ends with a Survey

This page looks identical to the first account creation page, to make the participants believe it is the same one. Users must fill in the same information as before, check the box, and click "submit". This time a popup appears thanking the user for the participation and that "I would appreciate it if you would follow the link to take a quick survey about your experience." This is followed by a button saying "Go to survey" redirecting the participant to a Qualtrix survey.

3.2.2.9. Survey – Qualtrix

Participants are sent to a standard Qualtrix survey accessed through the Utrecht University account. There they have 7 questions to answer, two where they rate their emotionality based on pictures on a 9-point scale, 4 where they rate their emotions on a word-based 7-point Likert scale, and finally one open-ended qualitative question. The survey is discussed in more detail in the subsection below.

3.2.3. Self-Assessment of Emotion (Survey)

The survey is a standard Qualtrix survey accessed through the Utrecht University account. The survey consists of 7 questions, two on a 9-point and four on a 7-point Linkert scale with one open-ended question. The first two questions are taken from the frequently used and researched Self-Assessment Manikin (SAM) scale. They ask participants to rate their feelings with images of human-like figures displaying varying degrees of emotion ranging from calm to aroused and distressed to pleased. These questions ask participants to rate their emotional response to the account creation process. The second question is inversed compared to the other two questions that inquire about valence, in the sense that it ranges from a negative emotion to a positive one. For that reason, the results from the second question were inversed before the analysis. The two questions have the following instructional text and images (Figure 3.1 and 3.2).

1. *"Arousal (low to high)*

   The first picture shows a person that is very calm or even sleepy. Relevant states could include tranquillity, relaxation, boredom, idleness or laziness. The last picture shows a person that is bursting with arousal. Relevant states could include excitation, rage, anger, euphoria, agitation or excitement. Please indicate how you felt after the "account creation" (not to be confused with the persona or character creation in the beginning of the experiment). Note that you can mark in between two figures."

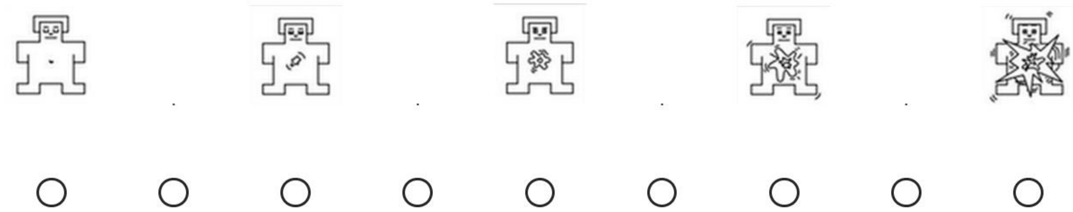


Figure 3.1. The five images signaling different levels of arousal for Question 1.

2. *"Valence (low to high)*
   The first picture shows a person that is distressed. Relevant experiences could include panic, despair, irritation or defeat. The last picture shows a person that is clearly pleased. Relevant experiences could include fun, happiness, satisfaction, delight or repose. Please indicate how you felt after the "account creation" process that you just completed (not to be confused with the persona or character creation in the beginning of the experiment). Note that you can mark in between two figures."

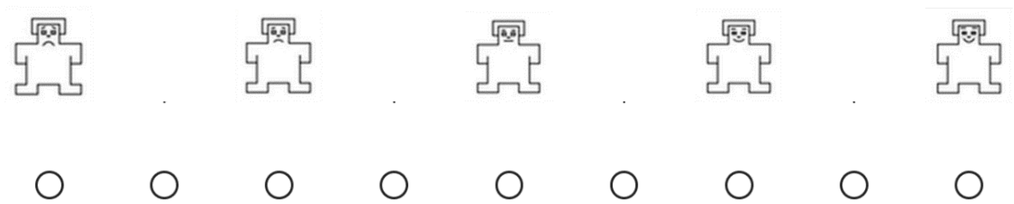


Figure 3.2. The five images signaling different levels of valence for Question 2.

The first two questions were supposed to use the strength of the SAM scale which it is supposed to be a quest way to assess emotions directly after they happen, and in this case the scale was presented immediately after the account creation task.
The following two questions ask about the character creation process on a seven point Likert scale. Both questions sounded the same:

"Please choose the option that best describes your state during the persona creation process in the beginning of the experiment"

The former question, asking about arousal, has the following options: *"completely calm", "mostly calm", "slightly calm", "neutral", "slightly anxious", "mostly anxious" and "completely anxious".* The second question, asking about valence, has the following options: : *"completely pleased", "mostly pleased", "slightly pleased", "neutral", "slightly annoyed", "mostly annoyed" and "completely annoyed.* The next two questions are identical to the previous two but ask about the current state of the participant:

"Please choose the option that best describes your current state"

The difference in wording is important here because the question is not asking about the emotionality towards the survey as is but the emotional state at the end of the experiment. The survey concludes with an open-ended question asking:

"Is there anything you would like to add, any suggestions or complaints?"

Participants are then thanked and that concludes the survey and the experiment. The survey as a whole can be seen in Appendix 3.

3.2.4. Breathing Measurement

Respiration will be measured by a breathing belt, biosignalsplux©piezoelectric respiration sensor (PZT: a respiration belt). A thoracic sensor is placed on the lateral side of the body slightly below the nipple line to get the highest peak-to-peak amplitude. Signals are measured in Volts and are recorded by OpenSignals (r)evolution software(Roes et al., 2022). OpenSignals is a software designed for biomedical signal processing and data visualization. It allows gathering and visualization of physiological data including respiration data. It was chosen for its simplicity, its robustness, and its compatibility with the breathing belt. The belt was connected to a laptop running OpenSignals and the software was configured to recognize the belt as a device measuring RESP (PZT). The sampling rate was set to 1000hz and baseline measures were taken in the form of having participants do a short session of "box breathing" or inhaling deeply and holding it for three seconds followed by an exhale with another three-second break until the next breath. The data gathered includes the timestamp of the beginning of the data tracking, the volts measured, and a row number. With 1000hz there are 1000 rows of data for every second. This data is then exported to a CSV file for further analysis.

3.2.5. Mouse Logging

To get an accurate timing for the tasks and the activity of the participant it was decided to track every mouse click performed on the computer during the experiment. This was done using AutoHotkey (AHK) an open source scripting language for Windows. AHK can keep track of keystrokes and mouse clicks and log them into a file. For this purpose, two scripts were created. The first one simply creates a CSV file and logs in each row the time of a click down to a millisecond, the button that was clicked (left/right), and the window where the mouse click happened. The second script creates a text file with each mouse click and stores the same data as the first one. The later script, however, deletes the previous text file before logging the click meaning that it only displays one row each time. This is done to take advantage of a feature of OSB studio, the screen capture software (which will be discussed later in this section), where the software can read a text file and display its content on the screen as soon as it is created. This allows for an easier task of getting the timing for the mouse clicks. This feature of the software has a problem where it sometimes fails to display the contents of the file if the clicks are too close to each other in time. For this reason, this dual script method was opted for ensuring all clicks are tracked even if they are not displayed on screen.

3.2.6. Recording Video

In addition to the mouse clicks being logged the whole experiment was recorded via a screen recorder (OBS Studio) that records everything happening on a pre-determined window, in this case Microsoft Edge where the website was running. This was done to be able to accurately track the time of each interaction with the interface and be able to accurately sync the breathing data with the actions of the participants. To do this it was decided to use Open Broadcaster Software (OBS) studios which is an open-source software used for video recording. OBS is an easy-to-use software with several good qualities. Firstly it allows the capture of a previously defined window making sure it will always only capture the video of the Microsoft Edge window containing the experiment. Secondly, OBS is highly flexible and allows for third-party plugins and scripts so it was possible to implement an overlay on the video tracking the time and the mouse clicks. The videos were recorded with a resolution of 1920x1080 and a framerate of 60 fps. There was no audio recording as it was not important for the purpose of the experiment.

3.2.7. Noting Video Times

The on-screen timer of the video turned out to lose its accuracy due to the limited 60 frames per second and latency. It was decided to add another software to get another measurement of the time to compare to the click log data. The solution was the program Avidemux which offers frame-per-frame forwarding with a time tracking down to a millisecond. Through this, it was possible to sync the starting time of the video with the starting time of the breathing data and get another method of pinpointing the time of task switching. This method is still limited to the 60 fps of the video which results in each frame spanning 16.67ms. but this offers an opportunity to compare the accuracy with the data from the mouse clicks. The times are then noted as time passed from the beginning of the video and are then converted to the current time using the time at the start of the video.

Times were noted in a document where video and click times were written down so they could simply be pasted into the Python notebook to be used to splice the data into different tasks. The first thing noted Is the length of the video, it is written down in the file to dispose of all breathing data that happens after recording has stopped. This is done to make the size of the dataset more manageable, the new dataset is referred to as the "trimmed" dataset. Having done that the video is fast forwarded to the begin of the study, when the submit button is clicked at the end of the consent form. For more details about the beginning and end of the tasks see the "Defining the tasks" section. When the moment is approached the first thing is to note the video runtime from the video, as that works the same for every participant. Following that, when everything is working as intended, the time of the last click performed after the rendering of the page following the button click appears on screen. This is then confirmed with the CSV file logging all the clicks as that time is the value used by default. This process has some problems, the two loggers sometimes have issues logging clicks that happen in sequence when the interval is particularly short. In the case one of them misses a log the other one can be used as a "backup". In the unlikely case that neither of them logs the click the time on screen is used, but that is less accurate due to the 60fp problem mentioned above. Having noted all the times any problems or incidents that should be removed are noted down in the document.

3.2.8. Difference Between Video and Click Logging

Video times were noted in using the click from the logs that were close to the action of the click, the last one to happen before it. In case of a double click, was investigated thoroughly using the video data and comparing between logs. There is an expected time difference between the click data and the video data since the video data is used to log the time when the button has its effect while the click data has the log time of the click. To accurately point out the frames used for the tasks it was the first frame of a greyed-out screen when there was a pop-up, the first frame of a white screen in case of redirect, and the graphics effects on the survey screen. This was the method opted for in the end.

3.2.9. Defining the Tasks

The experiment consists of four tasks, with the first one split into two task that are treated as individual tasks when it comes to breath analysis. The second subtask, however does not have a question in the survey, so there is no self-assessment data for that one. They vary in length and only the third task is created with the intent of provoking an emotion in the participants. Each task has two defined starting points and two endpoints. One is defined by a certain pre-determined frame in the video recording of the experiment and another one is defined by the exact moment of the click that leads to the beginning of the task, this timestamp is gotten from the two mouse log scripts. These two methods do not imply the same time, they are only recorded to get two perspectives. This will be described for each task in relation to the website.

1. The first task starts when the mouse click is performed on the button submitting the consent form. Using the video the task starts on the first frame of the following page, as soon as the screen turns from yellow to white. This task consists of participants reading through the directions of the experiment, creating the character, and browsing the website. This includes filling out the account creation form for the first time. The task ends the moment the submit

button in the account creation form is pressed. Using the video the task ends on the first frame after the screen gets a grey hue indicating a popup is coming. The first task was much longer and more diverse than the other two so it was decided to split it into two parts.

a. The first part consists of reading the instructions and creating a character. The task ends when the submit button is clicked from the mouse log and on the first frame of the next page for the video. This part will be referred to as "character creation" (Cc).
b. The second part covers all the browsing and the first account creation. It ends when the submit button for the account creation form is pushed for the mouse log and on the first frame with a grey hue from the video perspective. This will be referred to as the "shopping" task (S).

2. The third task starts at the exact moment the second one ends. The task begins with the first impossible-to-solve captcha followed by two similar captchas with red text below notifying the user that there is a limited amount of attempts left. After the third Captcha an error message pops up saying the information written on the form is wrong. After the participant has entered the information again the task ends the moment the submit button is clicked. From the video perspective, this task ends on the first grey frame following the click of the submit button.

3. The fourth task starts the exact moment the third one ends. The task consists of the Qualtrics survey questions. There are 7 questions, after they have been answered the task ends when the submit button is clicked. From the video perspective, the task ends with the first frame displaying the animation that follows the click of the submit button. This is indicated by the previously hollow yellow dots of the selected item for the survey becoming full.

The selection of the exact frames to begin and end each task is arbitrary but this was decided to make sure it was the same frame for each participant. The same can be said about the click logging, this was decided for the convenience and accuracy of the mouse click log. A thing to be considered about this method is that it simply starts when the other one ends, this means that all tasks begin with a delay of 0.5-2 seconds while the next task, or website, loads. This can be considered a problem. Adding the duration of the delay to the time of the click was considered an option but was dismissed due to the high and unpredictable variety in the length of the delays.

3.2.10. Participant Data

Each participant has a folder that holds the text file from the breathing data, the video recording of screen activity, and the mouse log file. Additionally, there is a text file with all the timestamps used for dividing the tasks and the whole Python workbook used to make their results. The processed files have four CSV files with breathing data and timestamps for mouse clicks based on the video, one for each task and one for the whole set. Another four with the same information based on the clicks. The results from the survey are included as a pdf file. Additionally, they have 2 CSV files with trimmed data where everything before and after the experiment has been cut. Finally, every subject has a normalized CSV file for every file already in their folder with all the data but normalized from 0-1.

3.2.11. Peak Detection

For detecting peaks and troughs in the breathing data the signal processing module of the SciPy library was used using the function scipy.signal.find_peaks version 1.11.4(*Scipy.Signal.Find_peaks — SciPy v1.11.4 Manual*, n.d.).The function is designed to identify peaks in a data sequence. It does this by finding the local maxima in a one-dimensional array. A peak is simply defined as a point that is higher than its neighbouring points and find_peaks identifies those points. The function's input is a one-dimensional array, just like the breathing data where the array holds a sequence of data points. It is possible to restrict and influence peak detection by various means and some of them will be mentioned below

The minimum height of the peaks can be specified using the height parameter. Using this the function only detects peaks above a certain value and is useful for breathing data that can have various fluctuations in between the breaths and troughs. Another useful restriction to avoid the detection of minor fluctuations between breaths is to set a minimum distance between peaks, this can be done using the distance parameter. In the current research, this is referred to as the expected breaths per minute and uses the fact that the breathing rate can be relatively stable over a short period. To get the best results from this parameter it is set to a fairly high expected breaths per minute that still has room for the more expected values. The two previously mentioned parameters are the ones currently in use in the data analysis. The threshold parameter allows for a restriction of the vertical distance between peaks. This can be useful for breath detection by eliminating the false positives on the downward slope from the peak that tends to happen in breathing data.

The appropriate values for the function were found using a slow trial and error method where different values were entered in the height threshold and distance between breaths and then the results were observed on the plots. There are more efficient ways of doing this and ways that would yield better results but this was deemed effective for the current research. There are instances where there are noticeable mistakes in the peak detection but they are few and far between. The values used for peak detection were 24 breaths per minute, which is objectively high. However, this is just the minimum distance between breaths, meaning that peaks detected before this distance will not

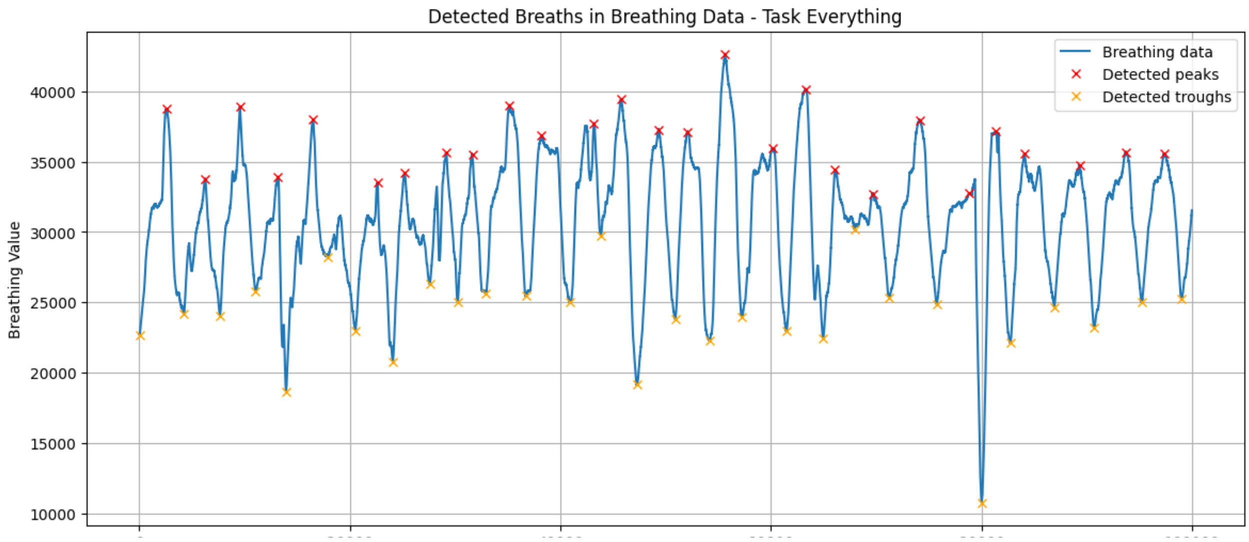


Figure 3.3. An example of a breathing graph with peaks and troughs marked on the image

be regarded as peaks, drastically decreasing false positives. The height requirement wound up being 0.55 quantile. This was inversed for the troughs and resulted in the most convincing graphs, an example of these graphs can be seen in figure 3.3 for an example of a small portion of a dataset and figure 3.4 for a whole set. It is consistently accurate with exceptions, like the one trough that can be seen around the 68,000ms mark. This peak detection can be optimized but it will take a long time and is something that would be good for future iterations of this study.

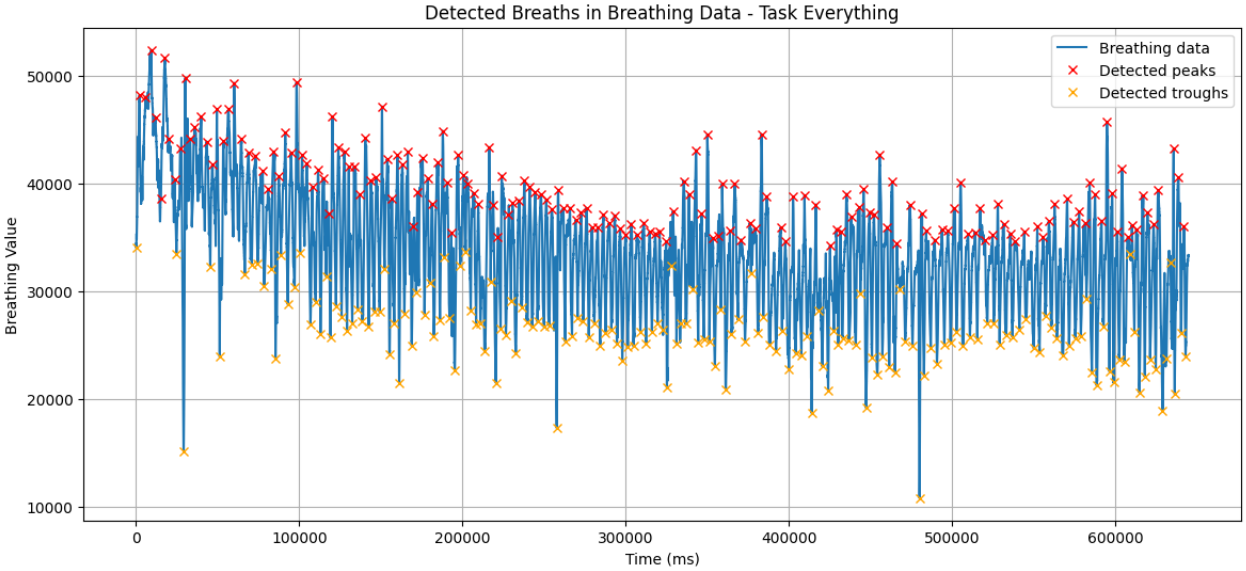


Figure 3.4. An example of a whole dataset with peaks and troughs marked on the image.

3.2.12. Regression Analysis

An ordinary least squares regression (OLS) analysis on the data was conducted to investigate how the previous task, and the breathing patterns associated with it, impacts or influences the detected respiration in the following tasks. This could give an idea of how emotion on a previous environment affects what happens in the next. OLS was chosen for its simplicity in use and in the interpretation of the results. I gives sufficient information for the depth of this study, but it will be interesting to apply more powerful models to this data. In this case simple linear regression is used to look at the pairs of breathing metrics and question. The OLS regression method was employed using the statsmodels.regression.linear_model.OLS class from the Statsmodels library in Python version 0.14.1(Josef Perktold et al., 2023). A weakness of this approach is that the experiment design can be thought of as repeated measures testing which implies a lack of independence between observations from the same participant. Independence of observations is not an assumption of the OLS regression, it however makes it more likely that there is dependence between the errors in the observations. It was decided to go forward, looking at each task as an independent observation, and inspect the results from the regression if it meets the assumptions. Various statistical tests were conducted to address the assumptions of the regression, followed by the inspection of plots. The results of these tests can be found in section 4.6 relationships between the tasks.

3.3. Pilot Studies

Before the main research, a series of pilot studies were conducted. The pilot studies were meant to give information about various aspects of the research. Firstly, they were expected to give an approximation of how long the experimental procedure would take. Second, the website itself had to be tested for bugs and usability. This did not have to be a robust inspection as the main body of the website was based on a premade template available within the Duda service. The shop itself was simply meant to act as a distraction for the rest of the website, the character creation and account creation which had to be more thoroughly tested as they were built for the experiment alone. Additionally, the pilot was supposed to assist with rephrasing and removal of bad or unclear survey questions and give the experimented some practice in managing the breathing belt and handling the data it gathered. The data gathered from the studies was used to tackle problems that might come up when working with the research data, figure out what software to use, and build a tool to process it. To satisfy these expectations seven preliminary studies were conducted with participants aged 25 to 60. They can be split up into three types of studies: online usability studies, in-person usability studies, and full pilot studies with breathing belts.

Three of the studies were online usability studies, these were expected to give qualitative information about the flow, uncover bugs, and get opinions on the tasks and how believable the complications at the end of the process are. Participants were sent a link to the website. They were asked to time themselves going through the process and follow the instructions of the experiment as they honestly would. Additionally, they were asked not to use the recommendation features of their browsers to automatically fill in the fields, name, email, password, etc. as this will be disabled in the main study. The three participants all took about 10 minutes to complete the task suggesting that they explored the website more than was required for the task. For the purpose of the pilot study answers to the field were logged and they showed that all participants honestly created their personas and did not tend to skip any fields or "cheat" in any way. The results were helpful, there were simple notes on design flaws such as some elements that jumped when scrolling down the website. The text with the instructions was unclear in some places and has since been clarified. The most important discovery from these usability studies was that there was an error in the flow between the sites where some purchase buttons did not redirect to the correct place and where the end of the study redirected to the home screen which led one participant to repeat the study multiple times before noticing. The notes on the account creation task were interesting where a participant said: "The letters to prove that I'm not a robot made me curse, even if I knew the purpose of the website". And another one said that "the captcha was mean" and "difficult to read". When asked about the level of frustration experienced by the task one replied that they "Did not get very frustrated with anything, it usually takes me some time to create an account".

The two in-person usability studies showed similar results with the added benefit of getting to watch the participants interact with the interface and using the computer that will be used to conduct the real study. These participants also got a survey in the end unlike the first three. When asked about how convincing the create account process appeared they did not say they had become suspicious even if it failed repeatedly and were more prone to blame themselves. These studies also revealed that it would be wise for the researcher to leave the participant alone for the duration of the experiment. They should stay out of sight to make the website experience more genuine as well as reduce the feeling of awkwardness when creating a persona and failing in the account creation.

The first full pilot study served as a training session on how to use the breathing belt, how to calibrate, and how to export the data. These studies gave interesting notes, that the character creation task could be too emotional or stressful for people that are not comfortable being asked to be creative on the spot. They were happy about the interface and were frustrated with the account creation process. One of the participants said that the captchas were impossible and that they awoke feeling that the participant had something wrong with them specifically. The most important aspect of these pilot studies was to gather the same type of data as in the main studies. The data included video data from the browser window, a table of logs for mouse clicks, breathing data in a table with 1000 logs for every second, and the survey answers. Since the respiration data needs to be highly accurate it immediately became a clear problem that the video and respiration had to be perfectly synced in order to capture the changes in breathing followed by events on screen.

4. Results

4.1. Questionnaire Results

The questionnaire consisted of seven questions, six quantitative and one qualitative. The first two questions were used from the SAM scale and were on a nine point scale and the following four questions were on a seven-point scale. The reason for the use of different scales was based on the idea that SAM has been used for a long time and is validated in its accuracy. It should be simple to understand and understandable for a varied age group. However, it has been noted that the SAM scale works best when it is used in concurrence or right after an emotional event. For that reason it was decided to only use the SAM scale to measure the emotional effect of the account creation process where things become difficult. Right after the task finishes participants are met by the scale, with the hope that the emotion is fresh in their memory, still ongoing. The following questions had a more standard 7-point Likert scale with neutral in the middle and extremes on each end. The reason for not repeating the SAM scale was firstly because the emotional response to the first question regarding the account creation process was prioritized and second that the time that had passed since the task would strip the SAM scale of its advantages.

The questions were split into three pairs of two questions, one meant to measure arousal and the other one valence. They will be referred to by the letters A and V following their task names respectively. The data gathered from these questions, for 19 participants, will be analyzed looking for trends that can be used as evidence for some difference in emotional responses to the different tasks. The responses from subject 18 were removed from the dataset since the breathing data was corrupt. Their mean response for this survey was within what was to be expected compared to others in the group.[2]

The questions related to arousal and valence will be compared separately. Every task has two questions assigned to it, one asking about valence and the other about arousal. Question two, asking about the account creation process had an inversed valence scale compared to the other questions on the questionnaire so the results from it were reversed, one became nine, two became eight, etc. Since the first two questions were on a 9-point scale it was decided to normalize them to a 7-point scale using equation 4.1 below:

$$\text{Normalized Value} = \left(\frac{\text{Original Value} - \text{Min Original Scale}}{\text{Max Original Scale} - \text{Min Original Scale}}\right) \times (\text{Max New Scale} - \text{Min New Scale}) + \text{Min New Scale}$$

Equation 4.1. Equation used to normalize the values in the first two question to for the 7-point scale

and round the result to four digits. The two account creation related questions had a more extreme response on average where they are higher than the others. The remaining four seem to be quite similar. The descriptive statistics can be seen in table A.1 in appendix 4. Question 1 is related to the account creation process, task 3, and measures arousal question 2 is also related to the account creation process, task 3, and measures valence. Question three is related to the character creation process, task 1 and measures arousal question 4 also relates to task 1 and measures valence. Question 5 is related the participant’s current state at the end of the experiment (survey), task 4, and measures valence. Finally, question 6 relates to the survey and measures arousal.

[2] More information on outliers can be seen in the *"outliers"* subsection in the *"methodology"* section.

4.1.1. Inspecting the Questions

To look at the questions on their own they were split into groups of arousal and valence related questions. Figure 4.1 shows boxplots for the six questions. It suggests that the scores tend to be higher for the first two questions related to account creation. The arousal at the end of the survey appears to be lower than in the other two tasks and with a relatively low variability.

The dataset is not big so the requirement of normal distribution is important to be able to perform an ANOVA to see if there is a difference in the means. The distribution of the six questions can be seen in figure 4.2. To get a better understanding of the distribution a Shapiro-Wilk test was conducted since it is an effective test for smaller sample sizes. The null hypothesis is that the sample is normally distributed. Table A.2 in appendix 4 has the results. Since the normalcy of half the questions can be disputed it was decided to perform a non-parametric tests on the data.

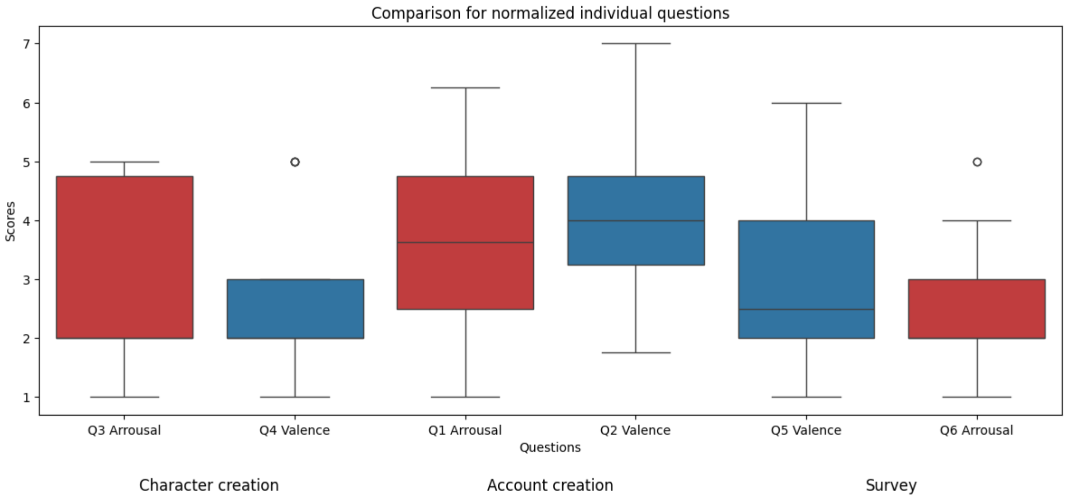


Figure 4.1. Boxplots of the six questions marked by task.

To compare the three groups a Kruskal-Wallis Test was conducted with the null hypothesis that the responses from all three tasks have the same median. A statistical difference in the medians of the three arousal responses (character creation, account creation, and survey) was found ($H(2) = 7.85$, $p = 0.020$) rejecting the null hypothesis and stating that there is a difference in the median between any two of the three pairs. To determine which pairs have a significant difference post hoc comparisons using the Mann-Whitney U test, another non-parametric test that assumes (H0) that the distributions of both groups are equal and are also well suited for non-normally distributed data. Since three pairs are being compared a Bonferroni correction was used to lower the risk of type 1 errors. There was a significant difference between the responses from the account creation (A) question and the survey question (A) ($U = 224.5$, $p = 0.047$) and account creation and character creation (A) ($U = 246$, $p = 0.008$) but not between the survey and character creation ($U = 146$, $p = 0.601$). The difference between account creation and the survey was to be expected since those were designed to be the most extreme and most neutral of the tasks. Character creation was also designed to be a relatively neutral task and the difference between account creation is to be expected, although not as severe as between character creation and the survey. The difference between character creation and the end of the survey is again not significant, even without the correction suggesting that the two tasks evoke similar levels of arousal.

For the valence responses, the Kruskal-Wallis Test suggested there was a statistical difference in the medians of the three arousal responses (character creation, account creation, and survey) (H(2) = 9.71, p= 0.008) rejecting the null hypothesis and stating that there is a difference in the median between any two of the three pairs. The results of the Mann-Whitney U test with Bonferroni correction show a significant difference in valence responses between account creation (V) and character creation (V) (U = 248, p = 0.019) and account creation and the survey (U = 243, p= 0.031), but not between character creation the survey (U = 150, 2.117). The significant difference between

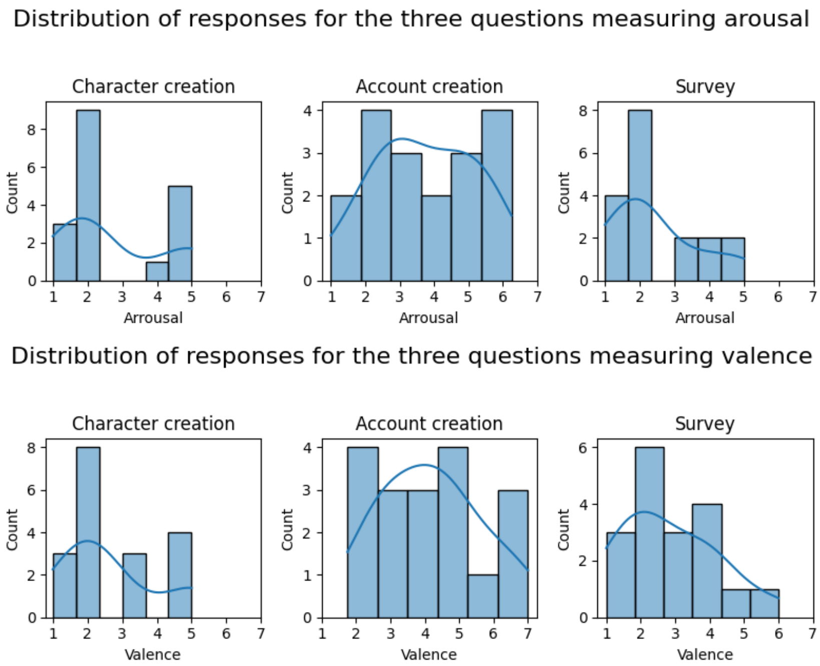


Figure 4.2. Distributions of responses for the questions by dimension.

account creation and the survey align well with the hypothesis that participants would be more annoyed or less pleased during the account creation when everything is problematic than after finishing the simple questionnaire at the end of the experiment. Similar to the arousal responses, there is a significant difference in self-reported valence between the account creation task and the two others. The lack of difference between character creation and the survey is also to be expected since both tasks were meant to be neutral although the character creation was expected to have more emotional load than the survey.

4.1.2. Relationships between the Questions

To get an idea of how the questions relate to one another, Figure 4.3 shows a heatmap with a correlation matrix. Noticeably, there is no correlation between arousal and valence on the account creation task so people seem to have different levels of anxiety and distress. There is in general little correlation between the account creation (V) question and any other question. The question asking about arousal during account creation significantly correlates with the question asking about the following survey task (A) r(16) = .57, p = .014. This suggests there could be an emotional carryover effect from the previous task on the survey task where the emotion experienced in the previous task colours the response on the survey task which was hypothesized to be neutral. Looking at figure 4.4

displaying the results from an ordinary least squares (OLS) regression analysis, the responses to the account creation (A) question explain a third of the variance in the responses to the survey (A) question, (F(1, 16) = 7.67, $p < .014$, $R^2 = .32$). The regression analysis model is discussed more thoroughly in the "Regression analysis" subsection 4.6.The fact that one task follows the other may explain a part of the similarity and predictability, but there is no other R^2 relationship between questions related to different tasks this high, suggesting there is a greater relationship there. Other noteworthy relationships are character creation (A) and its correlation with both character creation (V), r(16) = .50, p = .035 and The survey (V), r(16) = .46, p = .055. Although the latter correlation

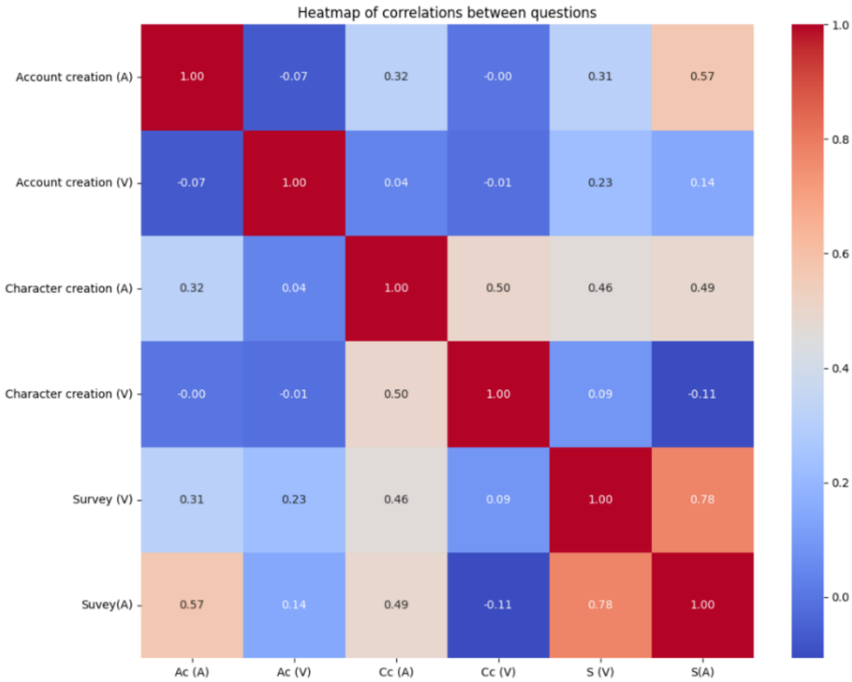


Figure 4.3. Heatmap with a correlation matrix for all six questions for character creation (Cc), account creation (Ac), and the survey (S).

is not significant to the standard of .05 it is only slightly above that. The results from Character creation (A) explain 25% and 21% of the variance of the answers to the character creation (V), (F(1, 16) = 5.31, $p < .035$, $R^2 = .25$) and survey (V) questions (F(1, 16) = 4.30, $p < .055$, $R^2 = .21$). This suggests that while the self-reported response from the character creation tasks has a less effect on the survey task than the account creation task did it is still significant, but interestingly only for arousal. While arousal predicts, at least partly the variation in responses to questions both related to arousal and valence there is no evidence of such prediction from the question about valence. Finally, there is a relatively high correlation between two questions asking about the survey task, r(16) = .60, p = 00 where the answer to the first explains more than half of the variation of the other ($F(1, 16) = 24.34$, $p < .00$, $R^2 = .55$). The same can be said about the arousal and valence questions for the character creation tasks as mentioned above, although not to the same extent. It is noteworthy that this relationship between arousal and valence questions is present in the tasks meant to be more emotionally light and not in the account creation task that is more emotionally heavy. Tables with both the F-values and the p-values for all combinations of questions can be found in Appendix 5**.**

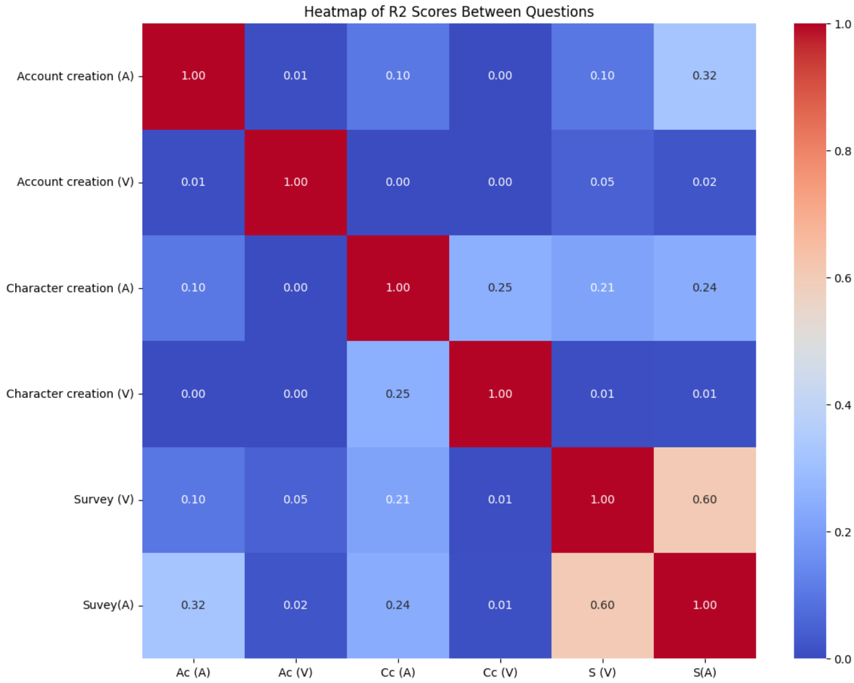


Figure 4.4. Heatmap with an R2 matrix for all 6 questions for character creation (Cc), account creation (Ac), and the survey (S).

Summary. The questionnaire, involving 19 participants, used seven questions to measure emotional responses to various tasks. It combined the SAM scale for the first two questions and a seven-point Likert scale for the others. The SAM scale was specifically employed to capture the immediate emotional impact of the account creation task. The questionnaire grouped questions to assess arousal and valence across different activities: account creation, character creation and a survey. The initial two questions, originally on a nine-point scale, were normalized to a seven-point scale for consistency. Statistical analyses, including Shapiro-Wilk, One-Way ANOVA, and Kruskal-Wallis tests, identified significant differences in arousal and valence responses among the tasks. Notably, the account creation task elicited more pronounced emotional reactions than the others. Additionally, the study uncovered other interesting facets where character creation and the survey tasks elicited varied but less pronounced emotional responses than account creation. A correlation analysis using a heatmap matrix indicated potential emotional carryover effects, particularly between the account creation and survey tasks. The results highlighted the varied emotional impacts of different tasks but drew interest to the heightened response to account creation.

4.2. Processing the Data for Breathing Analysis

The data gathered from the 19 participants was pre-processed, removing unnecessary data from the beginning and end of the experiment. The data was cleaned roughly, only ignoring extreme outliers. To further work with the data the peaks and troughs of the set had to be detected, these values can then be used to gather further information from the data. The data was normalized on a scale from zero to one to account for the individual differences in lung capacity. The purpose of this analysis was twofold, to compare the breathing data between the four tasks, and to look for correlations between the breathing data and the self-reported emotional states from the questionnaire. The tasks were compared on the following criteria: breaths per minute, the distance between breaths and throughs, relative volume of breaths, average depth of breaths, and sharpness. The variability of these values was also interesting since variations in breathing patterns might hold information on emotional states. The same values were then correlated with the results from the emotion questionnaire for each task.

4.2.1. Preprocessing of Data & the Syncing Problem

The breathing data was measured using OpenSignals (further discussed in the methodology section) thatching breathing at 1000hz. The output from OpenSignals was a large text file with 600.000-1.000.000 data points and metadata including the exact timestamp from when the recording was started. The text file was imported to a Python notebook, and metadata was cleaned leaving the timestamp. Since each datapoint accounts for one millisecond each datapoint could be given an exact timestamp. The experiment began with the participant putting on the breathing belt and starting the respiratory recordings, setting up the mouse tracking, and finally starting the screen recording. The screen recorder also has an accurate timestamp from the beginning of the recording, down to a millisecond. Having this information it was possible to mark the exact time of the beginning of the recording on the breathing data frame the text file from the mouse tracking was also imported and added to the data frame, so every click had a breathing data point along with it. The subsection "*noting the video times"* in the methodology section explains how the other timestamps were gathered, they were added to the data frame for each participant manually. This has the unfortunate problem of the frequency of the video recording being 60hz so due to the limitation of the technology the timestamps can be off by up to 16ms. Because the datasets usually consist of hundreds of thousands of data points that are then averaged it was decided to continue with this approach. It is however a possible point for improvement in a future iteration of the study. With all the timestamps in place, the data frame was cut into different sections: "everything", which was just the unadjusted data frame, "trimmed", which was the data frame after removing everything that happened before the experiment started (after the consent form was accepted), and then one for each task. This left each participant with six files of their data. The "everything" file would come in handy during the normalization process because it included the calibration period that usually happened before the screen recording was turned on.

4.2.2. Data Normalization

In the beginning of every experiment after fitting the breathing belt and before beginning the experiment, participants were asked to do a so called "box breathing exercise". This has them breathing in deeply, counting to three, and then exhaling until their lungs are empty, count to three and repeat. They were instructed to do that four to five times before beginning to read the consent form. This exercise has three advantages for the study. Firstly, it gets the subjects to relax and get in the moment, second it gives a rough estimation of their lounge capacity and third the distinctive pattern of taking a deep breath and counting to three, and so on is relatively easy to detect on a respiration graph. Knowing that the participants were box breathing until they started reading it was possible to estimate the time of their box breathing by looking at the time when they scrolled down from the first paragraph of the form to continue reading. This made it possible to narrow the window of data from the "everything" file down to around 30 seconds before the scrolling began and identify the box calibration there. Figure 4.5 shows an example of a visible box breathing period. Not that the peak detection algorithm does not work properly in a sample this small.

After locating the box breathing period that usually lasts around 30 seconds a function identified the highest peak and lowest trough of that period as an estimate of lounge capacity. With these numbers for each participant it was possible to normalize their data on a scale from 0 to 1 so everyone was on the same scale cutting out all higher values at 1 and lower values at 0. The formula used for normalization can be seen below in equation 4.2.

$$\text{Normalized\_Value} = \frac{\text{Value} - \text{trough\_value}}{\text{peak\_value} - \text{trough\_value}}$$

Equation 4.2. Normalization formula for the breathing data

All six files for every participant were normalized and saved separately. They were the only ones used from here on. This process did not come without problems and it was difficult to locate the calibration period for a few of the participants, it was however possible for everyone except for subject 18 (see outliers subsection). There was a problem with several participants where the beginning of the experiment appeared higher on the scale than the rest of the experiment, making it difficult to calibrate using the box breathing exercise, however, it was possible to locate a proper timeframe for most. Another approach was tested where, instead of a box breathing exercise, a period of deep breaths and long exhales was located in the whole dataset and used as a calibration period, finding the highest peak and the lowest trough in that period. It was decided to stick to the box breathing calibration period, but it would be interesting to see the difference between the two.

4.2.3. Outlier Removal

Outliers of 2 standard deviations above the mean were removed after inspection of the data. Outliers were removed from the following calculations: The height of peaks and troughs, and the distance between peaks and troughs. The removal affected the calculations of the sharpness of peaks and troughs in the way that those exceeding the cut-off limit were not considered for sharpness calculation. The decision was made after having studied the graphs of the respiration data noticing instances of high peaks or low troughs that were out of proportion to everything else. There were

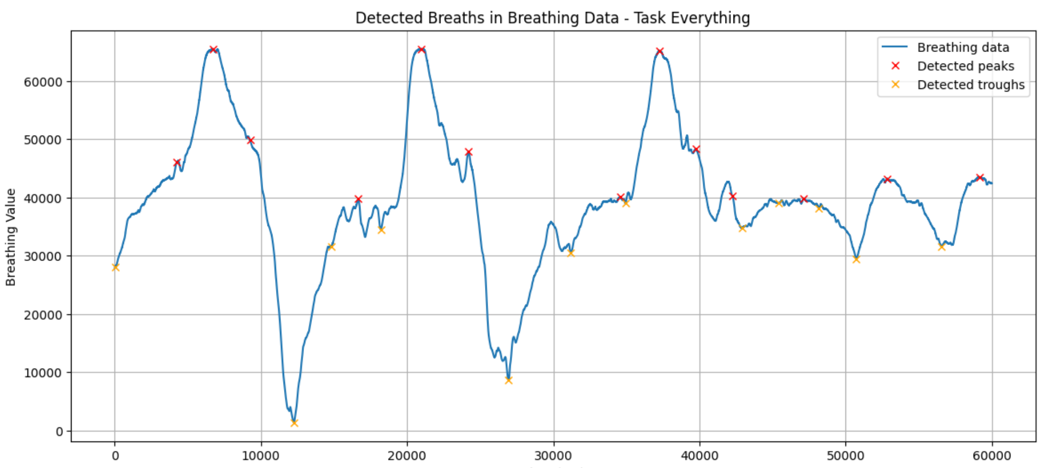


Figure 4.5. Box breathing on a respiration graph.

instances of large peaks without a fitting trough and vice versa. Having experimented with different methods, IQR method, and a higher multiple of standard deviations it was decided to remove those that exceed two standard deviations. The outliers meeting this criterion were relatively rare in these large datasets and tended to stick out in the charts as anomalies. It is however a possibility that some of the outliers removed were actual breathing events, due to the limitations of the study there is no way to investigate that after the fact. This could be mitigated by recording either sound or video during the experiment. Statistical tests were rerun on the datasets with outliers removed without their removal and without much change to the results or their significance increasing the confidence in the decision. More often than not the tests could not dispute normalcy.

The entire data from Subject 18 was removed from the dataset. This was due to an issue with the calibration. The calibration that took place at the beginning of the experiment was meant to assess the individual lung capacity and use that to normalize the data. In this case, all values in the first minute of the experiment far exceeded the rest of the data rendering the calibration period useless (figure 4.6). This effect is likely caused by a participant sitting upright during the beginning of the experiment and then slouching over the computer over the remainder of the experiment, especially if the breathing belt was tight around the chest.

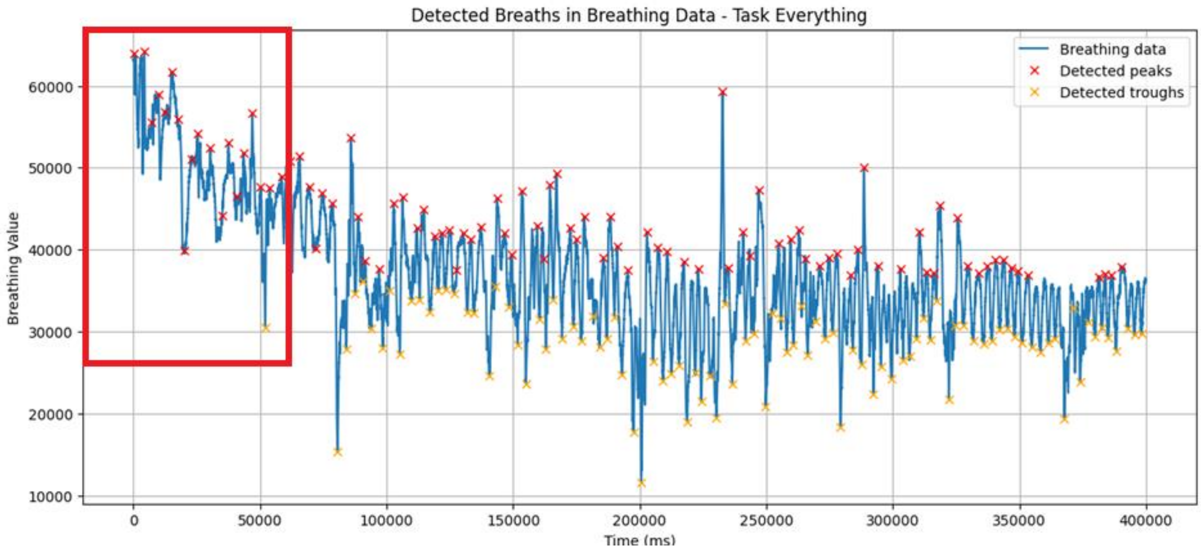


Figure 3.6. Respiration graph from subject 18 showing how the beginning of the recording is higher than the rest.

4.3. Comparing the Tasks

The breathing data was split up into four tasks, namely character creation, shopping, account creation, and the survey. The self-report questionnaire only asked about three tasks, with the first two tasks combined. The tasks were hypothesized to elicit different emotional responses. This hypothesis is partly supported by the results of the self-report results where there was a significant difference between responses to the tasks, especially between the third task and the fourth with the first one eliciting more emotion than expected. The following subsections will explore the differences between tasks to see if similarities can be seen in the breathing data. Below is a list of the tasks and their abbreviations. In each of the following subsections a Shapiro-Wilk test for normalcy was conducted because of the small sample size, along with tests for skewness and kurtoses. The results can be seen in appendix 4 along with descriptive statistics. Deviations from normalcy will be pointed out whenever observed along with what has been done to mitigate the deviations.

4.3.1. Respiration Rate

Breaths per minute are an indication of the respiration rate (RR). They are calculated by counting the number of peaks and troughs, exhales and inhales respectively, throughout the task and dividing the amount by 60. This results in five sets of inhales per minute and exhales per minute. By the nature of breathing, there should be a similar number of inhales and exhales. There is however some discrepancy in the data. A difference of two can be accounted for by the nature of the experiment where an inhale is performed before the task starts and then another one after it ends resulting in a difference of two. When the difference is higher than two it can be the result of an anomaly in the

$$\frac{(\text{peaks}/60) + (\text{troughs}/60)}{2} = \text{Respiration Rate}$$

Equation 4.3. A formula for the respiration rate using the mean of peaks and troughs per minute

breathing pattern where an inhale or an exhale is irregular, an instance of someone speaking, coughing, or an error in the peak detection. A table showing the differences between average inhales and exhales per minute for each task can be seen in Appendix 6 where the instances where the difference is above two have been marked in red. Due to the slight differences in these values, it was decided to use the average between the two as the indicator for breaths per minute called respiration rate (RR), see equation 4.3. The original values for inhales per minute (IPM) and exhales per minute (EPM) can be seen in Appendix 6.

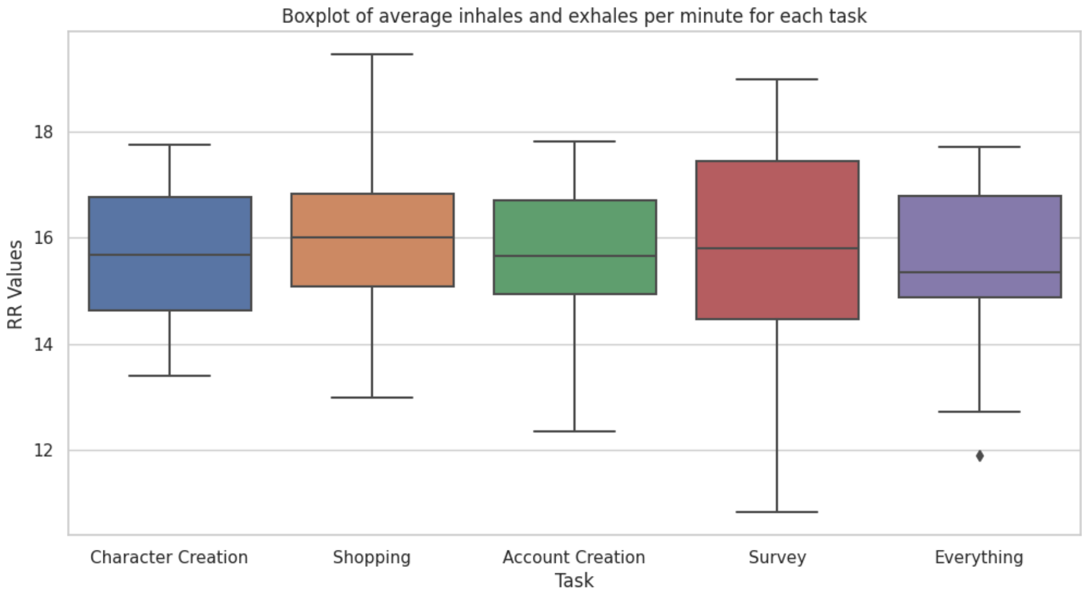


Figure 4.7. Boxplots of the average inhale and exhales per minute for each task and the experiment as a whole.

Figure 4.7 shows boxplots of the average respiration rate across the 18 participants for each task and the experiment as a whole. The average RR lies close to 16 breaths per minute for all the tasks resulting in approximately 3.75 seconds for each breath which is within the expected value for normal breathing (Badawy et al., 2017). The boxplots do not suggest there is a great difference in respiration rate between the tasks as is supported by table A.3 in appendix 4 where differences are low. There is more variation in the data from Task 4, which was the survey and was hypothesized to be neutral.

A one-way ANOVA conducted on the four tasks showed that it cannot be determined that respiration rate for any of the tasks differs significantly from the others $F_{4,14} = 0.21$, $p \leq 0.89$. Normalcy results can be found in table A.3 in appendix 4. To determine if any of the tasks are individually different from the combination of them all a non-parametric Wilcoxon Signed-Rank Test with Bonferroni correction was conducted on each task with the combination. The test indicated that only shopping was significantly different from the overall experimental data Z = 23.0, $p \leq 0.019$. Looking at the median in table 4.6 above it seems like the values for task 2, the shopping task, tend to be higher than for the overall experiment and is evidence for faster respiration during the shopping task. There does not appear to be a great difference between respiration rate between the tasks which could indicate that there is little physiological difference in emotion between them but the sample size is small so it is not easily generalizable.

4.3.2. Distance between Peaks and Troughs

The distance between peaks and troughs indicates the amount of time that passes between the most extreme of each inhale and exhale. These values distance between troughs (DPT) and distance between peaks (DBP) give an estimation of respiration rate and rhythm of respiration. The variability gives an idea of changes in breathing rhythm that could indicate that something is going on with the person breathing such as an emotional response. The distance between breaths can indicate how much air is inhaled and exhaled that can be indicative of shallow or deep breathing. Table A.5 in appendix 4 shows descriptive statistics for the average DBT and DBT and boxplots displaying the information in figure 4.8 with peak and trough distances displayed side by side for each task and the whole experiment.

Looking at the boxplots there seems to be a bigger variation in the distance between peaks than for troughs for all cases except for the account creation task, and can be seen clearly in the table as well. The median distance between breaths seems to be fairly stable across tasks with the exception of peaks in the first task. For the overall values there is a noticeable difference between the median of the troughs and peaks and a big one in the variability.

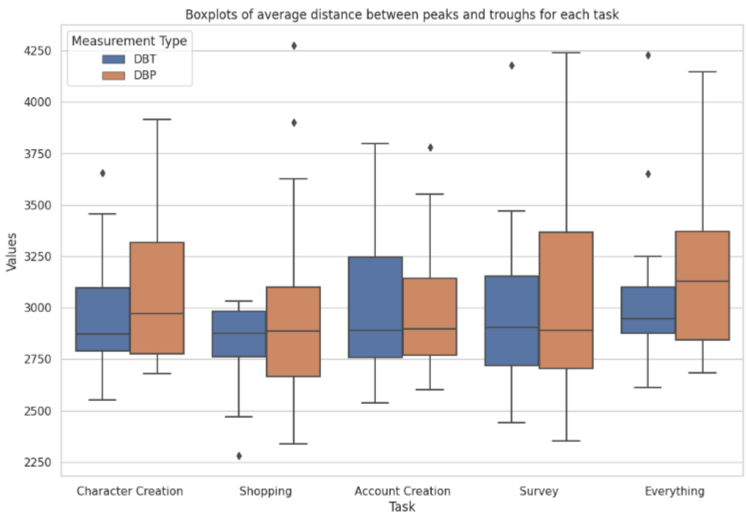


Figure 4.8. Boxplots of the average distance between troughs and peaks for each task and the experiment as a whole.

Having been able to dispute normalcy for almost half the results (table 4.6 in appendix 4) a non-parametric Kruskal-Wallis test on the four tasks to assess if they all have the same median for peaks and troughs, leaving the "everything" out as it is not independent of the other variables. The Kruskal-Wallis suggests there are no significant differences between the medians of distance between troughs for the four tasks ($H(3) = 1.09$, $p= 0.78$). The test also fails to show significant differences between the distances between peaks ($H(3) = 0.88$, $p= 0.83$). The test also fails to show significant differences if both peaks and troughs are taken together ($H(7) = 2.42$, $p= 0.93$). To see if any of the tasks differed from the entire experiment in terms of distance between peaks and troughs a Wilcoxon Signed-Rank Test with a Bonferroni correction was performed as it does not assume independence of the variables. character creation, account creation and the survey show no statistically significant difference from the whole experimental data but similar to the respiration rate (RR) shopping appears to be significantly different to the whole dataset both for troughs $Z = 22.0$, $p \leq 0.016$ and peaks $Z = 22.0$, $p \leq 0.016$. Respiration rate and distance between peaks and troughs are closely related and measuring the same thing so this does not come as a surprise.

The variability of breathing is something that could be effected by the task at hand or a change in affect. Changing from slow deep breathing to fast shallow breaths for example, or taking a few slow breaths in a time of frustration or focus. To compare the variance in distance between peaks and troughs between the four tasks a Levene's test was conducted with the null hypothesis (H0) that the variances are equal between them all. The test failed to show that there was a difference in the either DBP $F_{4,14} = 0.74$, $p \leq 0.53$ or DBT $F_{4,14} = 0.99$, $p \leq 0.40$. There does not appear to be a difference in the breathing variability between the tasks, at least not in the distance between the breaths, this can be due to a low number of participants or a a lack of difference between the tasks, or that the breathing variability only changes following more extreme emotional changes.

4.3.3. The Average Depth of Breaths

Calculating the average depth of breath can add to the information about breathing patterns. The depth or shortness of breath could inform on stress or intensity of a situation deeper breaths are, for example, often related to relaxation (Klausen et al., 2022). Here the focus is on inhaled air so the peaks are used for calculation. After removing peaks that are over two standard deviations from the mean the average is found and reported here. Table A.7 in appendix 4 shows the descriptive statistics for the average hight of the peaks (PH) for the 18 participants and figure 4.9 shows the boxplots for each task and the experiment as a whole.

There is a high similarity in the means with only character creation (task one) looking slightly higher than the others, the variance is similar with all cases. Looking at the median task three, the account creation task, has a visibly higher median on the boxplot and a difference of almost 0.04 from the next.

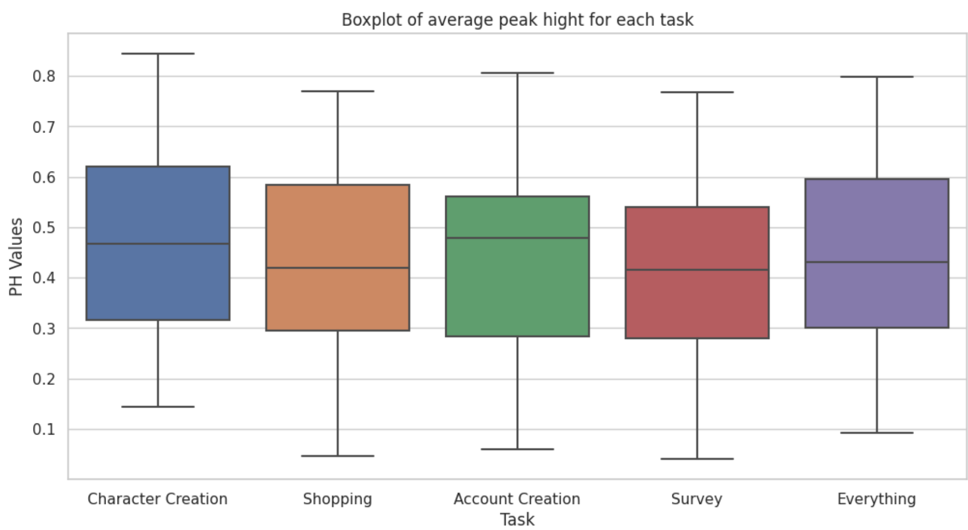


Figure 4.9. Boxplot of average hight of peaks for each task and the experiment as a whole

A one-way ANOVA was conducted (table A.7 in appendix 4 for normalcy tests) on the four tasks that showed that it can cannot be determined that the average height of peaks for any of the tasks differs significantly from the others $F_{4,14} = 0.43$, $p \leq 0.73$. To determine if any of the tasks are individually different from the combination of them all a non-parametric Wilcoxon Signed-Rank Test with Bonferroni correction was conducted on each task with the combination. The test indicated that character creation ($Z = 17.0$, $p \leq 0.006$), shopping ($Z = 12.0$, $p \leq 0.002$) and the survey ($Z = 6.0$, $p \leq 0.000$) differed from the data as a whole. These results suggest that this Wilcoxon Signed-Rank Test is surprisingly sensitive in this case, according to the results there is some slight difference in the hight of peaks in these three when it's compared to the whole time. There does not appear to be a significant difference between the peak hights between these tasks which could be a result of a small sample size or a lack of impact from the tasks. It is possible that much more is needed to impact the depth of breath so it is noticeable by these means.

4.3.4. Sharpness of Breath

Finally, sharpness of breath can be indicative of rapid and sharp breathing which is associated with anxiety or fear(Kreibig, 2010). This sharpness can both be measured on the inhale and exhale and will be estimated by assessing the steepness of the slope approaching each peak or trough. This measurement can be used to roughly estimate how rapidly inhales and exhales happen. To accomplish this a rolling window approach was used where a rolling window of 400 datapoints (ms). The pandas.DataFrame.rolling() function was used (*Pandas.DataFrame.Rolling — Pandas 2.1.4 Documentation*, n.d.) to calculate the slope within the window and then reporting the slope before each peak or trough. The window size of 400 was determined by experimentation and by the fact that the data set is so large. 400ms also fits got get the slope leading to a peak or trough without getting too much of the noise. This should work within a rhythm of normal breathing where the average breath cycle took 3.75 in this set. Table A.9 in appendix 4 shows the descriptive statistics for the average sharpness of peaks (PS) and troughs (TS) and figure 4.10 shows the boxplots for the four tasks and the experiment as a whole.

Looking at the boxplots the troughs are expectedly mirroring the peaks as the slopes are going in a negative direction towards them. The medians seem to be similar for both but the peaks during character creation seem to have a slightly higher median, especially higher than the survey, and a low variance except for outliers. The same effect of a greater sharpness during character creation can be seen for the troughs.

Having been able to dispute normalcy for almost half the results (table A.10 in appendix 4) a non-parametric Kruskal-Wallis test on the four tasks to assess if they all have the same median for peaks and troughs, leaving the "everything" out as it is not independent of the other variables. The Kruskal-Wallis suggests there are no significant differences between the sharpness of the peaks for the for the four tasks ($H(3) = 5.30$, $p= 0.15$). The test also fails to show significant differences between the sharpness of troughs ($H(3) = 3.45$, $p= 0.33$). To see if any of the tasks differed from the entire experiment in terms of distance between peaks and troughs a Wilcoxon Signed-Rank Test with a Bonferroni correction was performed as it does not assume independence of the variables. Character creation, shopping and account creation show no statistically significant difference in sharpness of peaks from the whole experimental data but the survey appears to be significantly different $Z = 9.0$, $p \leq 0.001$ for sharpness of troughs. Both peaks during character creation, $Z = 12.0$, $p \leq 0.002$ and troughs during survey $Z = 18.0$, $p \leq 0.008$ appear statistically different than for the data as a whole.

The survey was hypothesized to be the most normal and neutral task and in both cases does it have a lower median and mean for sharpness then the overall experiment suggesting that it might be more relaxing than the overall experiment. It is possible that people are more relaxed during the end of the experiment compared to the beginning where they have just went through the setup phase of getting the belt on and adjusting it.

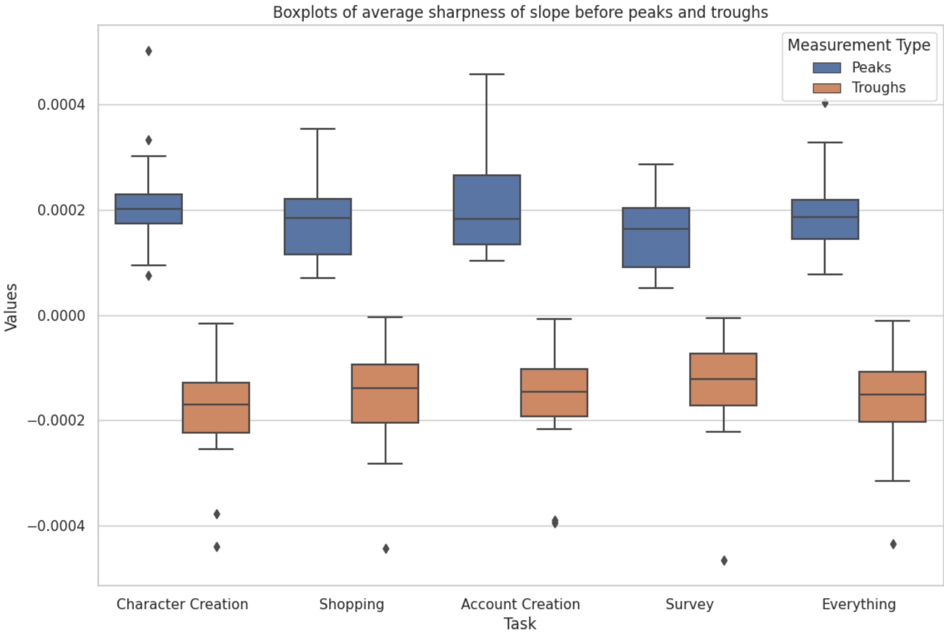


Figure 4.10. Boxplot of average sharpness of peaks and troughs for the tasks and the experiment as a whole.

Summary. The different comparisons of breathing did not show much difference between tasks when averaged over all 18 participants. There were instances of some of the tasks differing from the overall data, respiration rate (RR) was significantly different during shopping, the shopping task. Shopping differed again from the whole dataset when measuring the distance between both troughs and peaks, which is to be expected with the difference in respiration rate. Peaks were significantly lower during character creation, shopping and the survey than in the experiment overall while they did not differ for the account creation task. Task four had significantly different peaks and troughs from the rest of the data and looking at the descriptive statistics it seems to have milder peaks and troughs than the rest. Task one on the other hand, while also having different troughs does not have different peaks and seems to be sharper than the overall data suggesting that participants were breathing more harshly in the beginning of the experiment. The tasks do not seem to have been different enough to evoke a noticeable difference in breathing patterns, at least not for a small sample size. In the next subsection the respiration data will be correlated with the responses to the self-assessment questionnaire in the end of the experiment.

4.4. Comparing questions and Breathing Patterns

4.4.1. Account Creation – Arousal

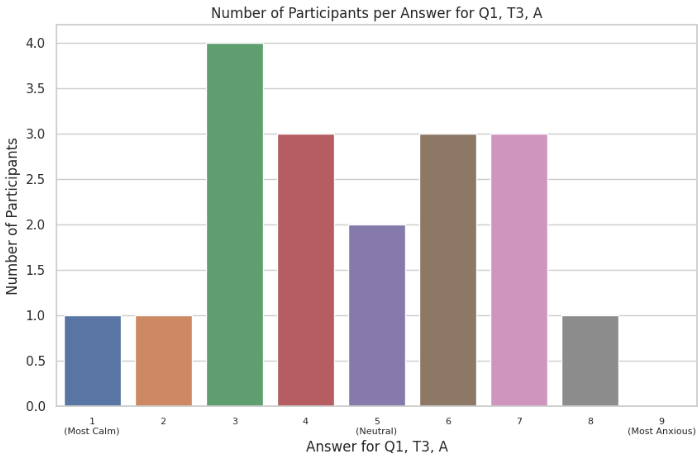


Figure 4.14 Distribution of answers for the questuion about the account creation task - arousal.

The question asked participants to rate their level of arousal during the account creation process. The account creation process was deliberately slow and designed to take more than one try to complete with problems on the way. Participants rated their level of arousal on a 9 point scale using images of a figure that was calm on the far left and gets visibly more aroused and anxious the further right on the scale it gets. The figure is depicted with a small shaking object in their stomach that grows bigger and bigger until it explodes. Due to the nature of the question it is not correct to talk about 5, the middle as neutral. It is just mid-way between "completely calm" and "bursting with arousal". Figure 4.14 shows the distribution of answers for the question. 2 people reported being completely or almost completely calm, 7 said they were mildly aroused to the task. 9 people were clearly aroused and bordering on anxious (a figure with a rather big shaking thing in their stomach in the first picture and another one where the thing is about to burst). This fairly even distribution suggests that the task was not stressful for all but still appears to have caused anxiety in a portion of the participants. The fact that some participants experience this as a rather calm experience could be due to the mundanity of the task, badly designed forms is something that people might encounter regularly and something that people have gotten used to.

	RR	**mDBT**	**mDBP**	**stdDBT**	**stdDBP**	**PH**	**PS**	**TS**
Account creation - A	-0.09	0.07	-0.01	-0.04	0.02	-0.27	0.36	0.13

Table 4.1. Account creation question - Arousal. Correlations with breathing data for the account creation task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.1 shows how the responses to the question correlate with the breathing data. None of the correlations are significant but it can be noted that there is a small negative correlation between the depth of breaths and the responses to the question $r(16) = -.27$, $p = .28$, suggesting that the breaths could be slightly shallower with people more anxious during this task which is in accordance with that would be expected. A full correlation table with all p-values can be seen in appendix 7**.** There is also a small positive correlation with peak sharpness $r(16) = .36$, $p = .14$, that would suggest that participants that rated themselves more anxious were taking sharper breaths during the account creation which fits well with them taking shorter breaths and also suggest a higher level of anxiety or stress. Respiration rate (RR) and distance between breaths do not seem to correlate with this question's results in any way, same can be said of the variation in distance between breaths.

Answer	**count**	**RR**	**mDBT**	**mDBP**	**stdDBT**	**stdDBP**	**PH**	**PS**[3]	**TS**
1.00	1.00	14.89	3313.29	2765.49	2221.85	1670.24	0.26	0.133	-0.127
2.00	1.00	17.80	2536.55	2777.55	1521.94	1708.28	0.81	0.227	-0.39
3.00	4.00	15.71	2861.70	3159.78	1774.99	2116.98	0.49	0.165	-0.176
4.00	3.00	15.77	2992.82	2890.11	1827.63	1762.86	0.56	0.178	-0.102
5.00	2.00	14.91	3134.94	3011.15	1920.93	1883.50	0.24	0.278	-0.159
6.00	3.00	14.33	3275.71	3292.77	2054.16	2327.33	0.32	0.195	-0.075
7.00	3.00	16.45	2842.48	2707.20	1733.40	1669.23	0.43	0.327	-0.207
8.00	1.00	15.37	2888.44	3052.64	1724.44	1904.31	0.33	0.162	-0.17
K-W p-value		0.47	0.38	0.34	0.54	0.23	0.33	0.5249	0.2224

Table 4.2. Average values for breathing data for each answer to question 1 – arousal, account creation, in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.2 shows the aggregated count for each answer to the question about the arousal level during account creation leaving out the options that has no answers. Due to the low number of answers and the nature of the data a non-parametric Kruskal-Wallis test was conducted on the 9 different measurements of breathing and the p-value can be seen in the bottom line of the table. It should be noted that due to the nature of the test the options with only one response were left out of the statistical test in order to keep its validity, the ones left out will be manually inspected. There were no significant differences found in the breathing signals depending on the answers to the questions according to the test. Looking at the averages there seems to be no trend in the respiration rate with the highest numbers being on answer two and seven. Same can be said of the distance between breaths and troughs. It is high for answers five and six which is about neutral. It has the highest value at six which could suggest a possible peak for the breathing parameter at this level of emotionality. There is still no clear trend to be found. The standard deviations of these measures indicate the variability in breathing patterns. These values do not show a clear trend with the answers relating to the account creation (A) answers, suggesting that variability in breathing does not significantly correlate with the arousal levels reported in this question. Peak hight or depth of breath has higher values on the lower end of the responses, with the exception of the one participant answering one. This fits with the mild correlation that was found in table 4.1 above and could suggest that deeper breaths during this tasks were linked to more calmness. The highest value of sharpness of peaks is at seven with three respondents, which would fit in with the correlation but there is no strong trend in the results.

Summary. While the task induced varying levels of arousal, the breathing data did not show clear trends or significant correlations with the arousal levels, except for some mild correlations in breath depth and peak sharpness. This suggests a complex relationship between physiological responses and perceived emotional states during this specific task.

[3] *Peak sharpness (PS3) and trough sharpness (TS) have been multiplied by 1000 for readability*

4.4.2. Account Creation – Valence

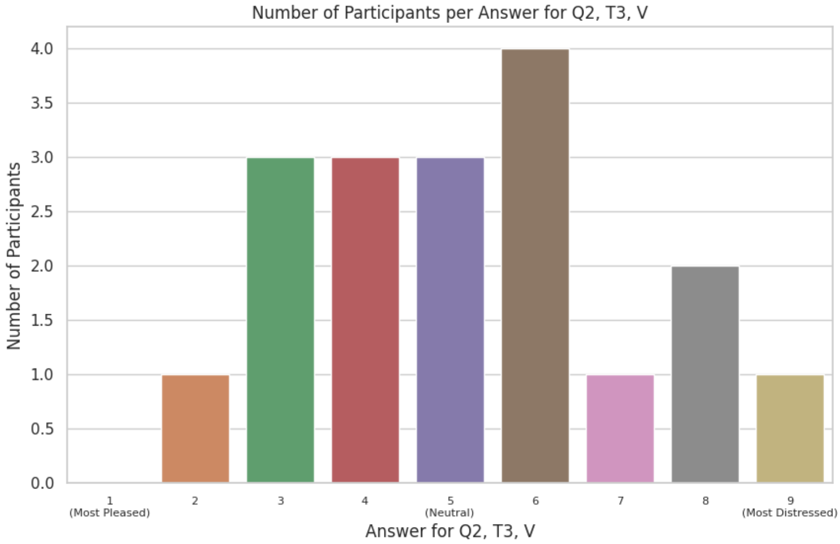


Figure 4.11. Distribution of answers for the question about account creation – valence

The question asked participants to rate their level of valence during the account creation process. Participants rated their level of arousal on a nine point scale using images of a figure that was pleased or happy on the far left and gets visibly less happy and has a neutral face on point five. The further right on the scale it gets the figure becomes more distressed with a big frown. Figure 4.11 shows the distribution of answers to the question. eight people report being more unhappy or distressed than neutral while seven are more pleased than neutral. Three people reported neutral feelings during the account creation process. There is a fairly even split between distressed and pleased although there is a longer left tail on the distribution so more people were highly distressed than were very pleased. Most people report being neutral or almost neutral to the task, with over half of the responses landing in the middle of the curve (10). While the feelings towards the task are mixed a surprising amount of people report being fairly pleased with the task, this could again be due to the mundanity of the task and that people have gotten used to problems with account creation and do not pay it much attention. Another possibility is that people answer more positively due to the fact that they are participating in an experiment. The task does not seem to have had a strong impact on self-assessed valence for the majority of participants.

	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
Account creation - V	0.14	-0.25	-0.04	-0.26	-0.11	-0.09	-0.43	0.08

Table 4.3. Account creation question - Valence. Correlations with breathing data for the account creation task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.3 shows how the responses to the question correlate with the breathing data. None of the correlations are significant but it can be noted that there is a medium negative correlation, that was close to significance, between the sharpness of peaks and the responses r(16) = -.43, p = .074. That would suggest that participants that rated themselves as unhappy were doing milder exhales breaths and the ones reporting being pleased were exhaling more sharply during the account creation. Respiration rate (RR) does not seem to correlate with the results from this question in any way, same can be said of distance between troughs and its variation, both having a low negative correlation r(16) = -.25, p = .30 and r(16) = -.26, p = .30 respectively.

Answer	Count	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS[4]	TS
2.00	1.00	16.72	2801.27	2879.66	1737.27	1783.33	0.77	0.419	-0.395
3.00	3.00	13.36	3489.66	3347.38	2250.90	2338.75	0.30	0.150	-0.095
4.00	3.00	16.73	2829.70	2706.29	1728.73	1686.62	0.45	0.342	-0.126
5.00	3.00	15.86	2867.71	3058.62	1712.30	1977.86	0.56	0.163	-0.159
6.00	4.00	15.85	2918.40	2858.32	1777.43	1777.24	0.34	0.222	-0.179
7.00	1.00	16.30	2683.36	2976.07	1651.89	1911.18	0.34	0.197	-0.199
8.00	2.00	15.36	3090.05	3124.40	1978.23	2039.29	0.53	0.104	-0.119
9.00	1.00	15.37	2888.44	3052.64	1724.44	1904.31	0.33	0.162	-0.170
K-W p-value		0.13	0.17	0.31	0.08	0.37	0.37	0.04	0.78

Table 4.4. Average values for breathing data for each answer to question 2, account creation, in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Looking at table 4 that shows the aggregated count for each answer to the account creation question about the valence level, leaving out the options that has no answers. the Kruskal-Wallis test found no significant differences in the breathing signals depending on the answers to the questions according to the test. Looking at the averages there seems to be no trend in the respiration rate. Same can be said of distance between breaths and troughs, although the highest number for both measures is for the answer three, on the pleased end of the curve. There is still no clear trend to be found. The standard deviations of these measures indicate the variability in breathing patterns. These values do not show a clear trend, suggesting that variability in breathing does not significantly correlate with the valence levels reported in this question. There appears to be a slight trend in the peak sharpness where there are higher numbers for the lower answers suggesting that participants that rated themselves as more pleased had sharper breaths, this is in tone with what the correlation table 4.3 above showed. The highest sharpness values are at two and four and they seem considerably higher than the values on the higher end of the answers.

Summary. The analysis of responses related to the valence level during account creation, reveals no significant differences in breathing patterns based on the Kruskal-Wallis test. The data shows an evenly distributed range of emotions from distress to pleasure, with no clear trend in respiration rate or distance between breaths and troughs across different valence levels. However, there's a slight tendency for sharper breaths in participants who reported feeling more pleased, particularly at lower valence levels (answers two and four), suggesting an inverse correlation between distress and breath sharpness. This pattern aligns with the observed medium negative correlation in the correlation table, although it's not statistically significant. Overall, the majority of participants reported neutral to slightly neutral feelings, with no strong impact on their self-assessed valence or varied breathing signals.

[4] Peak sharpness (PS3) and trough sharpness (TS) have been multiplied by 1000 for readability

4.4.3. Character Creation – Arousal

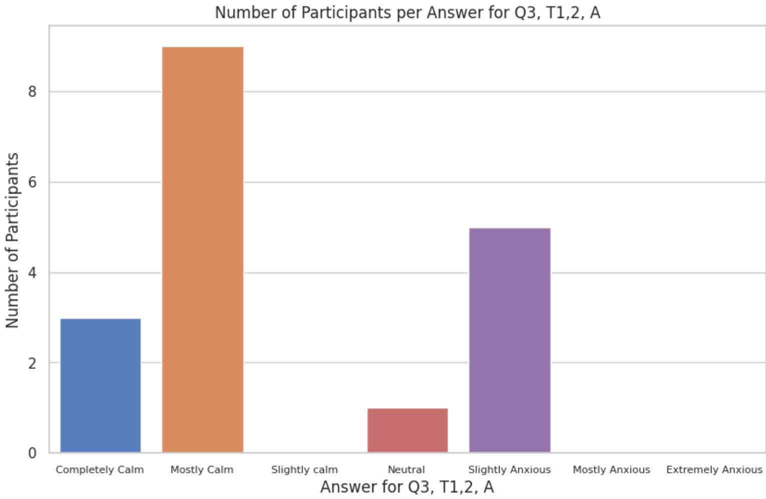


Figure 4.12. Distribution of answers for the question about character creation - arousal.

The question asked participants to rate their arousal level from the character creation process on a seven point Likert scale ranging from completely calm to extremely anxious. This task was the first one in the experiment so some time had passed from the completion of the task when the question was answered. Figure 4.12 shows the distribution of answers for the question. A clear majority were mostly or completely calm during the task (11) while five people were slightly anxious and one neutral. The slight anxiety experienced during this task matches with some of the qualitative feedback received during the experiment noting that some people can experience anxiety when put on the spot for some creative activity like creating a persona. This task was hypothesized to be a calm task. The shopping task had no question its relationship with this question are shown here as the two tasks were originally intended to be one and the same. They can now be thought of as two subtasks, character creation and shopping.

	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
Character Creation - A	-0.35	0.47*	0.28	0.47*	0.36	-0.24	0.4	-0.15
Shopping - A	0.05	-0.04	0.02	0.13	0.08	-0.2	0.46	-0.15

Table 4.5. Character creation question - arousal. Correlations with breathing data for the character creation and shopping tasks in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.5 shows how the responses to the question correlate with the breathing data for the character creation which refers directly to the question and the following shopping task, where participants were asked to go shopping as the character they created. Both the mean and variation of distance between troughs shows a significant correlation with reported valence during the character creation process, r(16) = .47, p = .0477 and r(16) = .47, p = .0483 respectively with the results from the regression analysis (F(1, 16) = 4.60, p < .0477, R^2 = .223) and (F(1, 16) = 4.57, p < .048, R^2 = .222) each explaining a fifth of the variance. This suggests a shorter time passed between inhales for the people that reported calmness during the character process and their breathing was more steady, while the people reporting more anxiety had longer intervals between inhales and displayed more variety in their breathing patterns. There is a small negative correlation with the respiration rate during the character creation r(16) = -.35, p = .157 but it is not significant. There is a medium correlation, that was close to significance, between the sharpness of peaks during both the character creation and shopping and the responses r(16) = .40, p = .099 and r(16) = .46, p = .057 respectively. That suggests that the participants reporting more anxiety during the task had sharper breaths during the task and they continued to be sharper throughout the shopping task. There is a mild correlation with the variation of peaks r(16) = .36, p = .14 for the character creation task which is to be expected since the variation of troughs was significantly correlated with increased anxiety. Other correlations were not significant and close to zero. The statistical tests for assumptions for the regression analysis results can be found in appendix 12.

Answer	Count	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS[5]	TS
1.00	3.00	16.34	2787.57	2889.04	1756.42	1837.11	0.53	0.194	-0.172
2.00	9.00	15.72	2916.23	3058.50	1843.05	2006.36	0.49	0.176	-0.170
4.00	1.00	16.80	2823.17	2775.79	1713.08	1732.20	0.84	0.502	-0.377
5.00	5.00	14.78	3180.49	3245.77	2294.14	2374.96	0.35	0.241	-0.186
K-W p-value		0.20	0.29	0.38	0.26	0.24	0.37	0.15	0.65

Table 4.6. Average values for breathing data for each answer to question 3 - arousal, character creation, in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Answer	Count	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
1.00	3.00	16.24	2781.94	2874.24	1718.49	1797.90	0.44	0.129	-0.151
2.00	9.00	15.76	2860.36	3059.51	1750.91	2067.35	0.44	0.158	-0.138
4.00	1.00	16.87	2935.87	2430.76	2059.55	1474.65	0.74	0.354	-0.232
5.00	5.00	16.06	2794.57	3068.99	1758.27	2118.59	0.31	0.209	-0.171
K-W p-value		0.78	0.44	0.85	0.81	0.80	0.45	0.29	0.56

Table 4.7. Average values for breathing data for each answer to question 3 - arousal, shopping.

Looking at tables 4.6 and 4.7 that show the aggregated count for each answer to the question about the arousal level during character creation (4.20) and shopping (4.21), leaving out the options that have no answers. the Kruskal-Wallis test found no significant differences in the breathing signals depending on the answers to the questions according to the test. Looking at the averages there seems to be no trend in the respiration rate. Distance between troughs shows a slight trend in the character creation task where the highest number (3180.49) is the commonly answered 5 which fits with the positive correlation pointed out in the correlation table 4.5 above. The ones that answered the highest had the longest mean distance while the ones that answered lowest had the lowest mean distance. A similar tend can be seen for the stdDBT. There does not seem to be noticeable trend in the shopping table with the exception of peak sharpness which is higher for answers 4 and 5, that trend is even more visible in the character creation table and fits with the medium correlations found.

Summary. the majority of participants reported feeling calm during the task, with fewer experiencing slight anxiety or neutrality. Correlation analysis revealed significant associations between reported calmness and certain breathing metrics: individuals who felt calmer had shorter intervals between breaths and more consistent breathing patterns. Although a negative correlation with respiration rate during the task was observed, it was not statistically significant. The data also suggested that participants who reported more anxiety had sharper and more varied breathing patterns. The Kruskal-Wallis test, however, found no significant differences in breathing signals across different arousal levels, aligning with the observed trends in breathing data. Notably, participants with higher anxiety showed longer mean distances between breaths and higher peak sharpness in both the character creation and subsequent shopping tasks, correlating with the qualitative feedback on anxiety during creative activities. Overall the task was pleasantly interpreted as hypothesized.

[5] *Peak sharpness (PS3) and trough sharpness (TS) have been multiplied by 1000 for readability*

4.4.4. Character Creation - Valence

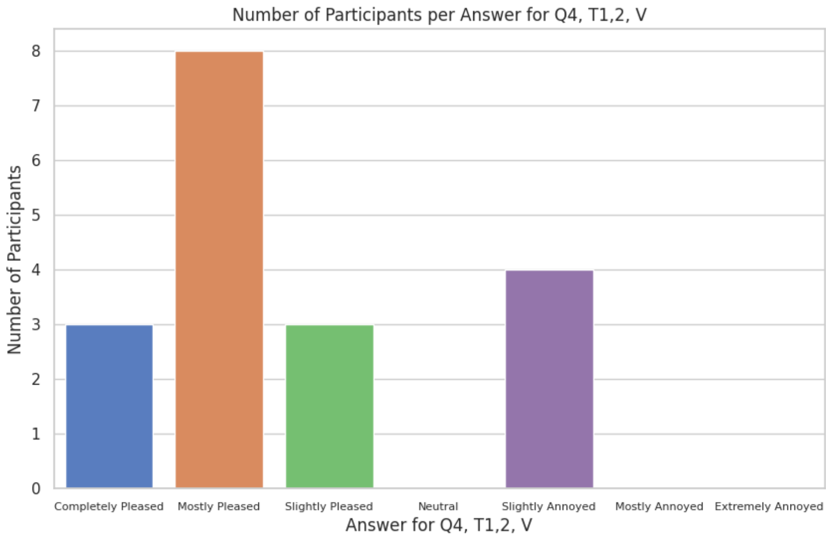


Figure 4.13. Distribution of answers for the question about character creation - valence.

The question asked participants to rate their valence level from the character creation process on a 7 point Likert scale ranging from completely pleased to extremely annoyed. This task was the first one in the experiment so some time had passed from the completion of the task when the question was answered. Figure 4.13 shows the distribution of answers for the question. A clear majority was slightly to completely pleased during the task (14) while four people were slightly annoyed. An even larger proportion of the participants report being pleased with the task than reported being calm about it in the previous question. There are still four people reporting to be annoyed a similar number that reported being and that was true of three out of four of them. The people reporting calmness were also quite pleased with the task so it is likely that the people that get anxious during creative creation similarly do not like it while the others do. As in the previous subsection this task can be split into two subtasks, character creation and shopping.

	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
Character creation - V	-0.27	0.47*[6]	-0.08	0.42	-0.02	0.07	0.02	-0.22
Shopping - V	0.14	-0.15	0.26	-0.21	0.34	0.1	0.19	-0.43

Table 4.8. Character creation question - valence. Correlations with breathing data for the character creation and shopping tasks in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.8 shows how the responses to the question correlate with the breathing data for the Character creation which refers directly to the question and the following shopping task, where participants were asked to go shopping as the character they created. There is a significant medium correlation between distance between troughs and the responses to the question r(16) = .47, p = .0474. This suggests that the people that were less pleased with the character creation had a longer pauses between their inhales, and similar to the question above about arousal there is a medium correlation between the variability of inhales as well r(16) = .42, p = .0799 although not quite

[6] * *if the p-value is ≤ 0.05 (significant at the 5% level),*
** *if the p-value is ≤ 0.01 (significant at the 1% level),*
*** *if the p-value is ≤ 0.001 (significant at the 0.1% level).*

significant. This fits well with the fact that more than half of those people that reported low valence also reported high arousal. There is a low negative correlation with breaths per minute for the character creation task r(16) = -.27, p = .283 again, mirroring the arousal responses. Interestingly there is no correlation with sharpness of peaks while there was a medium (although not significant) correlation with arousal. For the shopping task there is a negative medium correlation with the sharpness of troughs r(16) = -.43, p = .078 which is not quite significant but close. This would suggest that the people that were most pleased with the character creation task tended to inhale more sharply during the shopping experience. There is a low correlation between length between peaks (exhales) during the shopping task and responses to the question r(16) = .26, p = .296, r(16) = .34, p = .173. Other correlations are close to zero.

Answer	**Count**	**RR**	**mDBT**	**mDBP**	**stdDBT**	**stdDBP**	**PH**	**PS**[7]	**TS**
1.00	3.00	15.72	2767.96	3099.26	1692.89	2027.29	0.37	0.200	-0.161
2.00	8.00	16.01	2895.06	3035.85	1830.36	2007.57	0.47	0.230	-0.158
3.00	3.00	15.30	3092.42	3232.62	2195.74	2404.53	0.64	0.179	-0.256
5.00	4.00	15.00	3148.18	2978.95	2182.95	1954.89	0.44	0.225	-0.208
K-W p-value		0.65	0.25	0.89	0.21	0.93	0.43	0.99	0.49

Table 4.9. Average values for breathing data for each answer to question 4 - valence, character creation, in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Answer	**Count**	**RR**	**mDBT**	**mDBP**	**stdDBT**	**stdDBP**	**PH**	**PS**	**TS**
1.00	3.00	14.95	2940.92	3042.12	1847.09	1956.94	0.31	0.158	-0.108
2.00	8.00	16.15	2790.61	2891.27	1744.98	1892.83	0.41	0.184	-0.113
3.00	3.00	16.47	2910.54	2787.89	1826.86	1746.04	0.63	0.136	-0.229
5.00	4.00	16.06	2779.63	3328.46	1695.71	2453.97	0.38	0.214	-0.216
K-W p-value		0.72	0.66	0.67	0.62	0.57	0.22	0.63	0.25

Table 4.10. Average values for breathing data for each answer to question 4 - valence, shopping.

Looking at tables 4.9 and 4.10 that show the aggregated count for each answer to the question about the valence level during character creation (4.23) and shopping (4.24), leaving out the options that have no answers. the Kruskal-Wallis test found no significant differences in the breathing signals depending on the answers to the questions according to the test. Looking at the averages there seems to be only a slight trend with the respiration rate with lower people rating themselves more annoyed having slightly lower numbers (although not significantly lower). During the shopping task the respiration rate is noticeably the lowest for the people reporting the highest level of pleasantness (14.95). Distance between troughs shows a slight trend in the character creation task in the opposite direction, which fits the correlation from table 4.8 above. Otherwise there are no noticeable trends in the data.

Summary. Self-reported valence levels after a performing a character creation task, rated on a 7-point Likert scale showed that the majority (14 out of 18) reported being pleased, with only 4 slightly annoyed. This positive response was even more pronounced than the calmness reported in a related question. A significant correlation was observed between the participants' pleasure levels and their breathing patterns during the task: those less pleased had longer intervals between inhales and more variability in their breaths. While there was a negative correlation with breathing rate, it was not significant. The shopping task following character creation showed similar trends, with those who enjoyed the task inhaling more sharply. The Kruskal-Wallis test, however, found no significant

[7] Peak sharpness (PS3) and trough sharpness (TS) have been multiplied by 1000 for readability

differences in breathing signals across valence levels, with only slight trends observed in respiration rates.

4.4.5. Survey - Valence

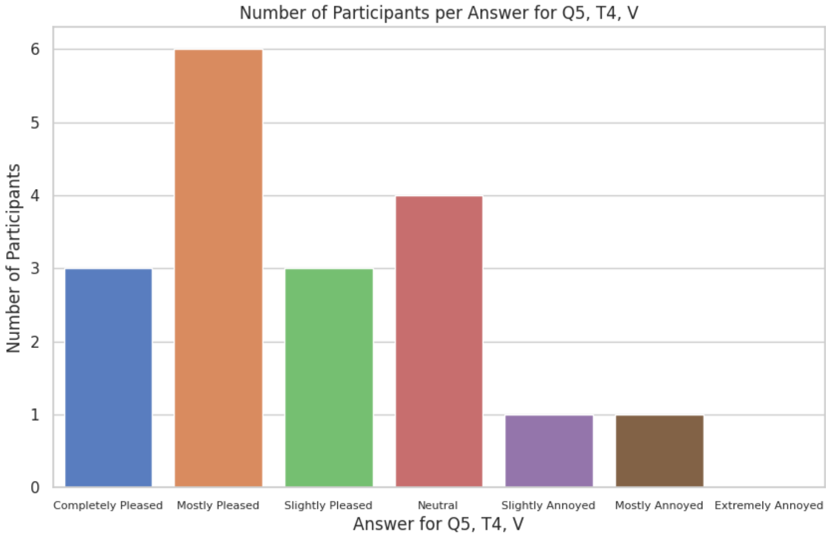


Figure 4.14. Distribution of answers for the question about end of the survey - valence.

The survey question asked participants to rate their valence level from the questionnaire process at the end of the experiment on a 7 point Likert scale ranging from completely pleased to extremely annoyed. This task was the last one in the experiment so there was an immediate feedback since participants were experiencing their emotions at the time the question was answered. Figure 4.14 shows the distribution of answers for the question. 12 participants rate their emotions as more pleased than neutral while only two are more annoyed. Four participants are neutral. The task was hypothesized to be neutral in nature but there is always the possibility that some emotion from the previous tasks spills over between the tasks, it should also be noted that this question is asked at the end of an experiment that usually takes between 12-14 minutes.

	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
Survey - V	0.58*[8]	-0.55*	-0.51*	-0.47*	-0.5*	0.14	0.29	-0.14

Table 4.11. Survey question - valence. Correlations with breathing data for the survey in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.11 shows how the responses to the question correlate with the breathing data for the questionnaire task. There is a significant mid-high correlation between breaths per minute and responses to the valence question $r(16) = .58$, $p = .0122$ where people reporting less pleasantness or more annoyance have higher breaths per minute. The regression analysis showed ($F(1, 16) = 7.97$, $p < .0122$, $R^2 = .33$) meaning RR explains a third of the variance of the responses to this question. This is expected since anger and annoyance have been connected to higher respiration rates (Kreibig, 2010). There is a significant negative correlation between distance between both peaks and troughs and the responses to the question $r(16) = -.55$, $p = .0184$ and $r(16) = -.51$, $p = .0316$ respectively with ($F(1, 16) = 6.89$, $p < .0183$, $R^2 = .30$) and ($F(1, 16) = 5.55$, $p < .0315$, $R^2 = .257$). This does not come as a surprise since it can be expected that a reduced distance between troughs and peaks (inhales and exhales) follows a higher respiration rate. This suggests

[8] * *if the p-value is ≤ 0.05 (significant at the 5% level),*
** *if the p-value is ≤ 0.01 (significant at the 1% level),*
*** *if the p-value is ≤ 0.001 (significant at the 0.1% level).*

that with increased pleasantness the interval between breaths increases and the breathing rate increases. Interestingly, there is also a significant negative correlation with the variability of both peaks and troughs $r(16) = -.55$, $p = .0498$ and $r(16) = -.51$, $p = .0348$ suggesting that with more annoyance the breathing gets more stable and rhythmic while there is more variability in the breathing patterns when a participant reports being pleased. These variables also held significant predictive power with ($F(1, 16) = 4.50$, $p < .050$, $R^2 = .220$) and ($F(1, 16) = 5.31$, $p < .035$, $R^2 = .249$) respectively. While not significant there is a low correlation with peak sharpness $r(16) = .29$, $p = .24$ with sharpness slightly increasing with annoyance which matches results from previous questions. There responses to this question correlate noticeably better with the breathing data than previous questions. This could be due to the fact that people are answering the question while the data is being recorded and therefore giving a more accurate description of their emotion. The statistical tests for assumptions for the regression analysis results can be found in appendix 12

Answer	Count	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS[9]	TS
1.00	3.00	15.29	3065.84	3055.50	1875.27	1998.16	0.46	0.099	-0.115
2.00	6.00	14.01	3268.14	3444.85	2101.37	2430.49	0.35	0.146	-0.141
3.00	3.00	15.59	2962.10	3010.73	1793.74	1924.71	0.36	0.155	-0.059
4.00	4.00	17.03	2693.70	2768.71	1623.34	1703.42	0.37	0.207	-0.189
5.00	1.00	17.94	2442.32	2352.84	1447.33	1403.59	0.68	0.231	-0.142
6.00	1.00	18.75	2682.19	2459.72	1628.44	1477.65	0.50	0.070	-0.146
K-W p-value		0.23	0.05	0.27	0.07	0.27	0.79	0.30	0.135

Table 4.12. Average values for breathing data for each answer to question 5 - valence, survey in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.12 shows the aggregated count for each answer about the valence level during the questionnaire leaving out the options that has no answers. The Kruskal-Wallis test showed significant difference between the average distance between troughs $p = 0.05$ meaning there is a difference between some of the breathing patterns depending on weather participants answered the question with completely pleased, mostly pleased, slightly pleased or normal. A post hoc Mann-Whitney U test was performed to assess which response groups differed and the participants that answered 2 (mostly pleased) differed from 4 (neutral) $U = 24.0$, $p \leq 0.0095$, with Bonferroni correction it is $p \leq 0.0571$ so this difference boarders on significance. The difference between the answer 2 (3268.14) and both 5 (2442.32) and 6 (2682.19), slightly annoyed and mostly annoyed respectively is even greater suggesting that there would be a significant difference between these means. 5 and 6 were however not a part of the Kruskal-Wallis test since both answers only had 1 participant behind them. But this would be interesting to investigate with a bigger number of participants. The difference in median standard deviation of distance between troughs was close to significance ($p = 0.07$) so a Whitney U test was performed and showed, again that there was a (close to significant) difference between answer groups 2 and 4 $U = 23.0$, $p \leq 0.019$, and with Bonferroni correction $p \leq 0.114$ suggesting that there was a difference in breathing variability between pleasantness and neutral. There is a clear trend in breaths per minute where participants giving higher answers breath faster with the invers being true about distance between peaks. Sharpness of peaks is highest around neutral, but has the lowest values on each extreme, the most pleased and most annoyed.

4.4.5.1. Survey - Valence and the Experiment as a Whole

The responses to this last question about valence had medium correlations with the breathing data of the whole experiment for some of the breathing measures. For all the other questions the results had low correlations with the combined tasks. This might suggest that the question asking about valence at “the current moment” is better suited to inform about the overall valence than the other

[9] Peak sharpness (PS3) and trough sharpness (TS) have been multiplied by 1000 for readability

questions asking participants to recall it. The correlation between the RR across the whole study and responses to the question were r(16) = .47, p = .0501 and overall RR predicts 22% of the variance of the responses (F(1, 16) = 4.48, p < .05, R2 = .219). Distance between troughs mirrors the result in inverse direction. And the variability of the distance between troughs r(16) = -.044, p = .065 showing that the variability decreases with annoyance. Overall variability of distance between troughs predicts 20% of the variance in the responses to the question (F(1, 16) = 3.936, p < .065, R2 = .197). Finally overall peak sharpness had a correlation with the responses, r(16) = .44, p = .068, and predicts 19% of the variance (F(1, 16) = 3.82, p < .068, R2 = .193). Full tables with the correlations with the overall data and the regression analysis results can be found in appendix 11.

Summary. Participants rated their valence level during a questionnaire on a 7-point Likert scale, significant correlations were found between their responses and various breathing parameters. Participants who reported less pleasantness or more annoyance exhibited higher breaths per minute, consistent with previous research linking annoyance to increased respiration rates. Additionally, a reduced distance between breath peaks and troughs, indicating a higher breathing rate, correlated with lower valence scores. The standard deviation of the distance between troughs also showed a near-significant difference, suggesting variability in breathing patterns correlates with emotional state. Overall, breathing data correlated better with self-reported emotions than previous questions, likely because responses were given concurrently with data recording.

4.4.6. Survey - Arousal

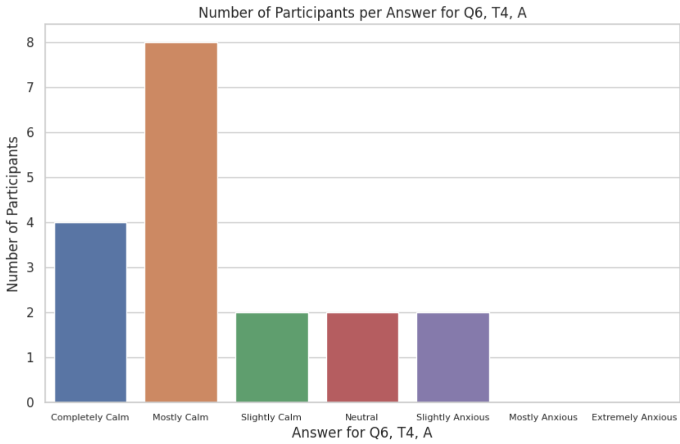


Figure 4.15. Distribution of answers for question 6 regarding the end of the survey- arousal.

The second survey question asked participants to rate their arousal level from the questionnaire process at the end of the experiment on a 7 point Likert scale ranging from completely calm to extremely anxious. Figure 4.15 shows the distribution of answers for the question. 14 participants rate their emotions as more calm than neutral while only 2 are more anxious. 2 participants are neutral. This task was hypothesized to be calm in nature so these results are expected. The 2 anxious participants had both rated themselves as slightly to mostly anxious throughout the experiment as a whole so their anxiety might have to do with something outside of the experiment.

	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
Survey - A	0.4	-0.36	-0.38	-0.26	-0.35	-0.1	0.24	0.1

Table 4.13. Question 6 - arousal. Correlations with breathing data for the survey in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Table 4.13 shows how the responses to the question correlate with the breathing data. None of the correlations are significant but it can be noted that there is a medium correlation with respiration rate, r(16) = .40, p = .104. That would suggest that participants that rated themselves as less calm were breathing slightly faster, following that there is a negative correlation with distance between both troughs and peaks r(16) = -.36, p = .137 and r(16) = -.38, p = .121 respectively. It is interesting that there are no significant correlations for this question since it is, just like the previous question responses were given concurrently with data recording. Participants reported more accurately on valence than arousal judging by that.

Answer	Count	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS[10]	TS
1.00	4.00	14.86	3155.18	3302.80	1962.57	2190.34	0.56	0.126	-0.194
2.00	8.00	15.34	2963.79	3081.81	1804.90	2056.15	0.34	0.144	-0.120
3.00	2.00	14.25	3419.31	3231.22	2308.22	2244.53	0.17	0.177	-0.015
4.00	2.00	18.46	2450.20	2375.11	1463.98	1443.21	0.54	0.258	-0.172
5.00	2.00	16.6	2760.97	2838.40	1675.24	1730.21	0.41	0.120	-0.150
K-W p-value		0.22	0.10	0.18	0.12	0.32	0.18	0.25	0.26

Table 4.14. Average values for breathing data for each answer to question 6 - arousal, survey.

Looking at table 4.14 that shows the aggregated count for each answer about the arousal level during the survey, leaving out the options that have no answers. the Kruskal-Wallis test found no significant differences in the breathing signals depending on the answers to the questions according to the test. Looking at the averages there seems to be a slight trend in the respiration rate with higher numbers following higher answer rating. The highest distance between troughs is found with the neutral answer(3419.31) but the lower answers tend towards higher means. There is still no clear trend to be found. The standard deviations of these measures indicate the variability in breathing patterns. There is a visible trend in both of the standard deviations where the variability of breathing seems to be higher when the participant is more calm, this has been seen with previous questions. There is no visible trend in sharpness or depth of breath.

For the calculations above it was decided to use the mean of the breathing data to aggregate the data by survey response. This means some nuance is lost and a possibility of significant differences going unnoticed when, for example finding the average respiration rate of the 9 people that answered question 3 with the option 2. To investigate whether the decision to use the means had an important influence on the results, the same procedure was repeated using the medians instead. The differences were minimal and it was not continued. The whole tables using medians instead of means can be seen in appendix 8 as well as the standard deviations of the means.

Summary. participants rated their arousal level during a questionnaire on a 7-point Likert scale from calm to anxious. Results showed most participants felt calmer than neutral, aligning with the task's hypothesized nature. Notably, the two anxious participants consistently reported anxiety, potentially unrelated to the experiment. Breathing data revealed no significant correlations with arousal ratings, though a medium correlation with respiration rate suggested less calm participants breathed faster. Additionally, a negative correlation was observed with the distance between breath peaks and troughs. Unlike valence ratings, arousal ratings showed no significant trends in breathing signals, indicating participants reported more accurately on valence than arousal.

[10] *Peak sharpness (PS3) and trough sharpness (TS) have been multiplied by 1000 for readability*

4.4.7. Summarized findings

The figure (figure 4.16) below shows the trends found in the relationships between the breathing data and the responses to the questions. it should be noted that not all of them are statistically significant to the $p \leq 0.05$ but some of them were closer to $p \leq 0.1$. The full table for tests for significance for the correlations can be found in appendix 7. As figure 4.16 shows there is a contradiction in the trends between character creation (the first task) and the survey (the last task) both for arousal and valence. This raises questions about whether participants answered more accurately when they were asked about their current state of emotion compared to 7-10 minutes ago. The results show that the breathing patterns measured during the survey better fit the literature than the ones observed during the character creation(Kreibig, 2010; Siddiqui et al., 2021).

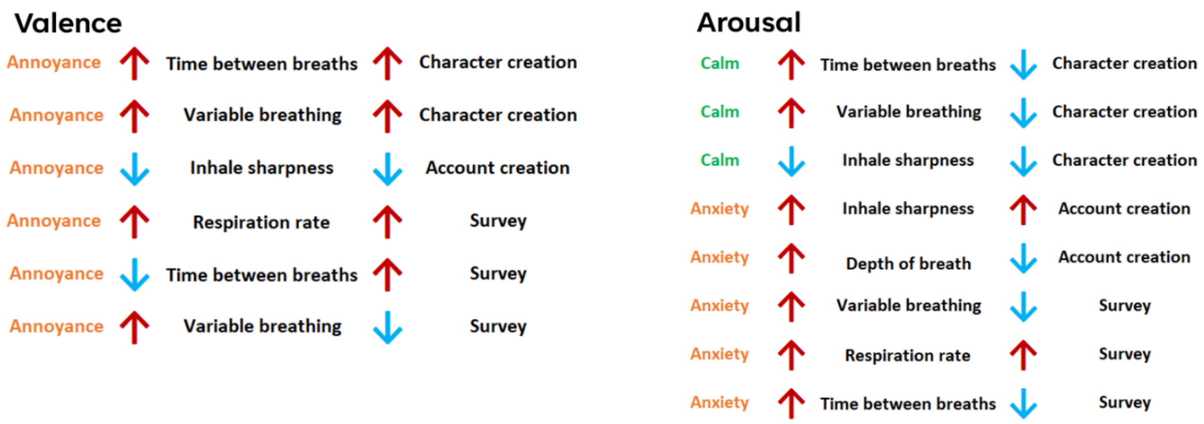


Figure 4.16. An infographic showing how the observed breathing patterns relate to the responses to the self-assessment questions at the end of the experiment for valence and arousal. The data is sorted by the tasks where the breathing patterns were observed and their respective questions.

4.4.8. Correlation of Breathing Patterns with Aggregated Arousal and Valence Scores

To find out if there was a relationship between overall answers to the questions about valence and arousal and the average respiration measurements for the entire duration of the study the responses to the 3 questions were summed up. The sums for the participants were then correlated with their overall breathing measures. The only significant correlations found were a correlation between overall arousal score and peak sharpness $r(16) = .531$, $p = 0.023$ and an inverse correlation between overall valence score and the variability of the distance between breaths $r(16) = -.439$, $p = 0.068$ with the latter one only being close to being significant for p-value < 0.05.

4.5. Exploring Task-to-Task Relationships through Breathing Patterns

An ordinary least squares regression (OLS) analysis on the data was conducted to investigate how the previous task, and the breathing patterns associated with it, impacts or influences the detected respiration in the following tasks. This subsection will introduce a correlation matrix followed by an R^2 matrix for each double of tasks to give an idea of how each task influences the breathing patterns of the next. While the vast majority of variable couples meet the assumptions of the analysis the exceptions should be kept in mind when interpreting the results where they are not as robust as the rest. The results from the statistical tests for all couples can be seen in appendix 12 where all the potentially problematic results have been coloured red.

4.5.1. *Characte Creation and Shopping*

Cc S	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.48*[11]	-0.16	-0.57*	-0.12	-0.51*	-0.01	0.06	-0.08
mDBT	0	-0.07	0.23	-0.05	0.23	0.06	0.1	-0.14
mDBP	-0.47	0.23	0.38	0.23	0.32	-0.07	-0.14	0.21
stdDBT	0.07	-0.1	0.14	-0.04	0.15	0.06	0.07	-0.15
stdDBP	-0.34	0.21	0.27	0.27	0.24	-0.05	-0.16	0.14
PH	-0.03	0.19	-0.01	0.22	-0.02	0.97***	-0.02	-0.57*
PS	0.56*	-0.33	-0.58*	-0.02	-0.51*	0.09	0.85***	-0.26
TS	-0.09	-0.27	0.14	-0.31	0.1	-0.68**	-0.35	0.89***

Table 4.15. Correlation matrix of the breathing measures from the character creation (Cc) and shopping (S) task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Looking at the correlation table 4.15 between the character creation task and the shopping task there is a significant correlation between the breaths per minute from task to task $r(16) = .48$, $p = .044$ suggesting that there is some similarity in the respiration rate between the tasks. Full matrices for all correlations with p- values can found in appendix 9**.** Table 4.16 below shows that the respiration rate during the character creation task explains a significant amount of the variance in the respiration rate on the shopping task ($F(1, 16) = 4.77$, $p < .044$, $R^2 = .23$). This is a relatively limited but significant effect on the respiration rate in the later task considering they happen in the span of around 5 minutes. There is a highly significant high correlation between peak height, peak sharpness and trough sharpness between the two tasks, $r(16) = .97$, $p = .00$, $r(16) = .85$, $p = .00$ and $r(16) = .89$, $p = .00$ respectively. The predictive power of these variables is expectedly high as well, ($F(1, 16) = 299$, $p < .00$, $R^2 = .949$), ($F(1, 16) = 49.9$, $p < .00$, $R^2 = .728$) and ($F(1, 16) = 62.4$, $p < .00$, $R^2 = .796$). This suggests that there is a close relationship between peak hight, and peak and trough sharpness between the two tasks and they do not seem to change a lot between them. Looking exclusively at the pairings of the same type of breathing measures there are no more significant relationships between the two tasks. The effect the character creation task has on the shopping task seems to be small so the two do not have much in common when it comes to respiration. This could suggests a change in breathing patterns when participants do the shopping task when compared to the character creation task. Full tables of F statistics and p values for every combination can be found in Appendix 9**.**

[11]* *if the p-value is ≤ 0.05 (significant at the 5% level),*
** *if the p-value is ≤ 0.01 (significant at the 1% level),*
*** *if the p-value is ≤ 0.001 (significant at the 0.1% level).*

Cc S	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.230*[12]	0.00	0.22	0.01	0.12	0.00	0.317*	0.01
mDBT	0.03	0.01	0.05	0.01	0.05	0.04	0.11	0.07
mDBP	0.326*	0.05	0.14	0.02	0.07	0.00	0.338*	0.02
stdDBT	0.01	0.00	0.05	0.00	0.07	0.05	0.00	0.10
stdDBP	0.260*	0.06	0.10	0.02	0.06	0.00	0.261*	0.01
PH	0.00	0.00	0.01	0.00	0.00	0.949***	0.01	0.469**
PS	0.00	0.01	0.02	0.00	0.03	0.00	0.728***	0.12
TS	0.01	0.02	0.04	0.02	0.02	0.330*	0.07	0.796***

Table 4.16. A matrix of the results from the (OLS) regression analysis for the character creation (Cc) and the shopping (S) task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

4.5.2. Character Creation and Account Creation

Ac Cc	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.57*	-0.67**	-0.45	-0.70**	-0.43	0.04	0.15	-0.3
mDBT	0.01	0.1	-0.1	0.18	-0.1	0.1	0.11	0.06
mDBP	-0.48*	0.65**	0.32	0.67**	0.29	-0.1	-0.28	0.38
stdDBT	0.08	0.03	-0.17	0.09	-0.15	0.11	0.11	0.07
stdDBP	-0.29	0.45	0.14	0.48*	0.13	-0.07	-0.28	0.34
PH	0.16	-0.23	0.23	-0.18	0.19	0.94***	0.13	-0.66**
PS	0.45	-0.3	-0.42	-0.3	-0.38	0.16	0.75***	-0.52*
TS	-0.29	0.33	-0.01	0.31	0.04	-0.68**	-0.34	0.89***

Table 4.17. Correlation matrix of the breathing measures from the character creation and the account creation task.

Looking at the correlation table 4.17 between the character creation task and the account creation task there is, again, a significant correlation between the breaths per minute from task to task $r(16) = .57$, $p = .0133$ suggesting that there is some similarity in the respiration rate between the tasks. It is noteworthy that the correlation is higher between these tasks than the previous tasks suggesting they have more in common. Both tasks involve filling out a form and writing while the shopping task was focused on exploration. Table 4.18 below shows that the respiration rate during the character creation task explains a significant amount of the variance in the respiration rate on the account creation task ($F(1, 16) = 7.75$, $p < .0133$, $R^2 = .33$). This is a relatively limited but significant effect on the respiration rate in the later task and again it explains more of the variance than it did for the shopping task. There is a highly significant high correlation between peak height, peak sharpness and trough sharpness between the two tasks, $r(16) = .94$, $p = .00$, $r(16) = .75$, $p = .00$ and $r(16) = .89$, $p = .00$ respectively. The predictive power of these variables is high as well, ($F(1, 16) = 125$, $p < .00$, $R^2 = .887$), ($F(1, 16) = 20.6$, $p < .00$, $R^2 = .563$) and ($F(1, 16) = 61.3$, $p < .00$, $R^2 = .793$). This suggests that there is a close relationship between peak hight, and peak and trough sharpness between the two tasks and they do not seem to change a lot between them. There is a noticeable difference in the R^2 for peak height compared to the previous comparison. The correlation and predictive power of peak height, peak sharpness and trough sharpness are relatively constant between all doubles suggesting that it does not change drastically over the course of the experiment. Full for F and p values and tables can be accessed in appendix 9. Looking exclusively at the pairings of the same type of breathing measures there are no more significant relationships between the two

[12] * *if the p-value is ≤ 0.05 (significant at the 5% level),*
** *if the p-value is ≤ 0.01 (significant at the 1% level),*
*** *if the p-value is ≤ 0.001 (significant at the 0.1% level).*

tasks. The effect the character creation task has on the account creation task seems to be relatively small. The two tasks do not follow each other so it is understandable that the patterns do not line up although the tasks share some similarities. It is possible that the first task has very distinct breathing patterns compared to the rest simply because it is the first task. It could be due to excitement, curiosity or the fact that they are beginning an experiment.

Cc **Ac**	**RR**	**mDBT**	**mDBP**	**stdDBT**	**stdDBP**	**PH**	**PS**	**TS**
RR	0.326*[13]	0.00	0.233*	0.01	0.09	0.03	0.20	0.08
mDBT	0.445**	0.01	0.417**	0.00	0.21	0.05	0.09	0.11
mDBP	0.21	0.01	0.10	0.03	0.02	0.06	0.18	0.00
stdDBT	0.488**	0.03	0.445**	0.01	0.229*	0.03	0.09	0.10
stdDBP	0.19	0.01	0.08	0.02	0.02	0.04	0.14	0.00
PH	0.00	0.01	0.01	0.01	0.01	0.887***	0.02	0.467**
PS	0.02	0.01	0.08	0.01	0.08	0.02	0.563***	0.11
TS	0.09	0.00	0.14	0.01	0.11	0.434**	0.266*	0.793***

Table 4.18. A matrix of the results from the (OLS) regression analysis for the character creation and the account creation task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

[13] ** if the p-value is ≤ 0.05 (significant at the 5% level),*
*** if the p-value is ≤ 0.01 (significant at the 1% level),*
**** if the p-value is ≤ 0.001 (significant at the 0.1% level).*

4.5.3. Character Creation and Survey

S	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
Cc								
RR	0.13	-0.34	-0.3	-0.35	-0.26	0.01	0.19	-0.18
mDBT	0.4	-0.25	-0.17	-0.23	-0.19	0.12	-0.05	-0.04
mDBP	-0.2	0.4	0.34	0.43	0.3	-0.08	-0.31	0.27
stdDBT	0.45	-0.28	-0.26	-0.24	-0.24	0.12	-0.07	-0.07
stdDBP	-0.04	0.26	0.15	0.3	0.16	-0.04	-0.33	0.19
PH	0.16	-0.11	-0.22	-0.18	-0.34	0.94***[14]	-0.09	-0.54*
PS	0.60**	-0.55*	-0.65**	-0.44	-0.55*	0.07	0.69**	-0.12
TS1	-0.4	0.29	0.44	0.3	0.46	-0.67**	-0.45	0.83***

Table 4.19. Correlation matrix of the breathing measures from the character creation and the survey task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

The two tasks furthest away from each other in time, the character creation task and the survey, are also different in nature. The character creation task asks for creative input and writing while the survey simply has the participant read and answer simple questions. There are no significant breathing pattern correlations between the two task with the exception of the aforementioned three, peak height, peak sharpness and trough sharpness that are high and highly significant , $r(16) = .94$, $p = .00$, $r(16) = .69$, $p = .001$ and $r(16) = .83$, $p = .00$ respectively. The same can be said about the predictive power of the variables, there is none at all anymore for respiration rate nor any other variable, with the exception of the three that are noticeably lower than in the past two comparisons PH ($F(1, 16) = 120$, $p < .00$, $R2 = .882$), PS ($F(1, 16) = 14.8$, $p < .001$, $R2 = .481$) and TS ($F(1, 16) = 34.2$, $p < .00$, $R2 = .681$), especially peak sharpness explaining less than half of the variance for peak sharpness of the survey task. It is to be expected that the predictive power of the variables reduces as time passes and other tasks come in between.

Cc	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
S								
RR	0.02	0.16	0.04	0.20	0.00	0.03	0.360**	0.16
mDBT	0.11	0.06	0.16	0.08	0.07	0.01	0.301*	0.08
mDBP	0.09	0.03	0.11	0.07	0.02	0.05	0.419**	0.19
stdDBT	0.12	0.05	0.19	0.06	0.09	0.03	0.19	0.09
stdDBP	0.07	0.03	0.09	0.06	0.03	0.12	0.303*	0.21
PH	0.00	0.01	0.01	0.02	0.00	0.882***	0.01	0.450**
PS	0.04	0.00	0.10	0.01	0.11	0.01	0.481**	0.21
TS	0.03	0.00	0.07	0.00	0.04	0.296*	0.02	0.681***

Table 4.20. matrix of the results from the (OLS) regression analysis for character creation and the survey task.

[14] * *if the p-value is ≤ 0.05 (significant at the 5% level),*
** *if the p-value is ≤ 0.01 (significant at the 1% level),*
*** *if the p-value is ≤ 0.001 (significant at the 0.1% level).*

4.5.4. Shopping and Account Creation

Acs	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.83***[15]	-0.70**	-0.75***	-0.72***	-0.72***	0.19	0.54*	-0.31
mDBT	-0.46	0.31	0.39	0.32	0.34	0.02	-0.43	-0.05
mDBP	-0.73***	0.57*	0.77***	0.64**	0.76***	-0.18	-0.49*	0.29
stdDBT	-0.33	0.25	0.26	0.27	0.24	0.06	-0.3	-0.14
stdDBP	-0.62**	0.46	0.69**	0.54*	0.70**	-0.18	-0.43	0.25
PH	0.28	-0.32	0.14	-0.26	0.09	0.98***	0.14	-0.66**
PS	0.3	-0.14	-0.35	-0.16	-0.35	0.08	0.89***	-0.47*
TS	-0.39	0.43	0.14	0.41	0.17	-0.62**	-0.38	0.85***

Table 4.21. Correlation matrix of the breathing measures from the shopping task and the account creation task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

The account creation task flows straight from the shopping task without any notice for the participant. They are just prompted to fill out a form, this could imply that breathing should be connected between the two tasks. The account creation task begins to act difficult and that could influence breathing patterns. Looking at table 4.21 the breathing in the two tasks is much more correlated than with previous comparisons. Respiration rate in the two tasks is highly significantly correlated $r(16) = .83$, $p = .00$ and the same can be said about distance between peaks $r(16) = .77$, $p = .00$ and its standard deviation $r(16) = .7$, $p = .001$. PH, and TS remain high and fairly stable $r(16) = .98$, $p = .00$, $r(16) = .85$, $p = .00$ and peak sharpness has regained its strength $r(16) = .89$, $p = 00$ and seems to be more sensitive for the temporal distance than the other two. Respiration rate during the shopping task predicts a large part of the variance of respiration rate in the account creation task ($F(1, 16) = 35.2$, $p < .00$, $R^2 = .687$) and the same can be said about distance between breaths and it's standard deviation ($F(1, 16) = 22.8$, $p < .00$, $R^2 = .587$) and ($F(1, 16) = 15.4$, $p < .001$, $R^2 = .491$) respectively, although to a lesser degree. The same is again true for the three last variables with peak height explaining almost all the variance in the later task ($F(1, 16) = 379$, $p < .00$, $R^2 = .96$) but peak and trout sharpness are high as well ($F(1, 16) = 61.7$, $p < .00$, $R^2 = .794$), ($F(1, 16) = 42.9$, $p < .00$, $R^2 = .728$) It is clear that there is more relationship between these two tasks than what has been seen until now. It is possible that the breathing patterns established during the shopping experience just continue with minimal change during the experience even if there are some problems with the form, that is just taken as a part of the experience. It is also possible that the issues with the form were just too normal and mundane for the participants. The seamless flow between the tasks could also have an impact.

Acs	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.687***	0.485**	0.567***	0.518***	0.521***	0.04	0.291*	0.10
mDBT	0.22	0.10	0.15	0.10	0.11	0.00	0.19	0.00
mDBP	0.529***	0.331*	0.587***	0.406**	0.582***	0.03	0.241*	0.09
stdDBT	0.11	0.06	0.07	0.08	0.06	0.00	0.09	0.02
stdDBP	0.391**	0.22	0.472**	0.290*	0.491**	0.03	0.19	0.06
PH	0.08	0.10	0.02	0.07	0.01	0.960***	0.02	0.438**
PS	0.09	0.02	0.13	0.03	0.12	0.01	0.794***	0.221*
TS	0.16	0.19	0.02	0.17	0.03	0.390**	0.15	0.728***

Table 4.22. A matrix of the results from the (OLS) regression analysis for the shopping and the account creation task.

[15] * *if the p-value is ≤ 0.05 (significant at the 5% level),*
** *if the p-value is ≤ 0.01 (significant at the 1% level),*
*** *if the p-value is ≤ 0.001 (significant at the 0.1% level).*

4.5.5. Shopping and Survey

EoS S	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.78***[16]	-0.79***	-0.79***	-0.75***	-0.77***	0.15	0.47	-0.25
mDBT	-0.51*	0.45	0.46	0.4	0.47*	0.05	-0.31	-0.11
mDBP	-0.54*	0.54*	0.65**	0.49*	0.65**	-0.12	-0.39	0.21
stdDBT	-0.39	0.36	0.31	0.36	0.38	0.08	-0.23	-0.04
stdDBP	-0.45	0.46	0.55*	0.41	0.57*	-0.13	-0.35	0.18
PH	0.27	-0.2	-0.3	-0.27	-0.45	0.99***	-0.09	-0.57*
PS	0.52*	-0.46	-0.54*	-0.37	-0.47	-0.01	0.84***	-0.15
TS	-0.48*	0.39	0.47	0.4	0.49*	-0.63**	-0.54*	0.95***

Table 4.23. Correlation matrix of the breathing measures from the shopping task and the survey task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

It can be seen in tables 4.23 and 4.24 that there is a clear relationship between the breathing patterns during the shopping task and the later survey task. The same variables have a medium to high correlation as in the previous comparison between shopping and account creation although to a lesser extent. This could be the same effect seen in the first four comparisons where the relationship diminished the further the tasks were apart in time. Respiration rate is significantly correlated between the tasks r(16) = .78, p = .00 and explains more than half the variance of respiration rate in the survey task ($F(1, 16) = 25.2$, $p < .00$, $R^2 = .612$). Distance between peaks and its variability are significantly correlated r(16) = .65, p = .003, r(16) = .57, p = .0132 and explain a less than half ($F(1, 16) = 11.9$, $p < .003$, $R^2 = .427$) and a third ($F(1, 16) = 7.77$, $p < .0132$, $R^2 = .327$) of the variance of the measures in the survey task respectively. Peak hight, peak sharpness and trough sharpness are very highly correlated again r(16) = .99, p = .00, r(16) = .84, p = .00, r(16) = .95, p = .00 respectively. The three variables also have high predictive power: PH ($F(1, 16) = 557$, $p < .00$, $R^2 = .972$), PS, ($F(1, 16) = 39.1$, $p < .00$, $R^2 = .709$) and TS ($F(1, 16) = 150$, $p < .00$, $R^2 = .904$) again explaining most of the variance of the same measures in the survey task, with trough sharpness with TS notably higher as compared to the previous relationship with account creation explaining 90% of the variance. This contradicts the hypothesis that the temporal gap lowers the predictive power between variables. 4

EoS S	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.612***	0.618***	0.620***	0.557***	0.586***	0.02	0.22	0.06
mDBT	0.261*	0.20	0.21	0.16	0.221*	0.00	0.10	0.01
mDBP	0.288*	0.294*	0.427**	0.244*	0.419**	0.02	0.15	0.05
stdDBT	0.15	0.13	0.10	0.13	0.14	0.01	0.05	0.00
stdDBP	0.20	0.21	0.299*	0.17	0.327*	0.02	0.13	0.03
PH	0.07	0.04	0.09	0.07	0.20	0.972***	0.01	0.327*
PS	0.275*	0.22	0.286*	0.14	0.22	0.00	0.709***	0.02
TS	0.235*	0.15	0.22	0.16	0.241*	0.395**	0.293*	0.904***

Table 4.24. The matrix of the results from the (OLS) regression analysis for the shopping and the survey task.

[16] * *if the p-value is ≤ 0.05 (significant at the 5% level),*
** *if the p-value is ≤ 0.01 (significant at the 1% level),*
*** *if the p-value is ≤ 0.001 (significant at the 0.1% level).*

4.5.6. Account Creation and Survey

The final double, account creation and the survey are adjacent tasks so it would be expected that they have more in common than tasks further apart. It should be noted that according to the survey results the emotional weight of the account creation task was rated significantly higher than the other task making inviting the possibility of emotional carryover effects into the following task. Looking at the tables 4.25 and 4.26 it can be seen that all variables are significantly correlated between the two tasks, most of them medium or highly correlated. They also all have a significant predictive power over their counterparts in the following task. Respiration rate has a correlation of r(16) = .74, p = .00 between tasks and explains more than half the variance ($F(1, 16) = 19.4$, $p < .0004$, $R^2 = .549$). Both the distance between troughs r(16) = .72, p = .0008 and peaks r(16) = .65, p = .004 have a medium correlation and explain close to half the variance in the following

S Ac	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
RR	0.74***[17]	-0.75***	-0.80***	-0.73***	-0.78***	0.36	0.31	-0.47
mDBT	-0.68**	0.72***	0.75***	0.73***	0.75***	-0.38	-0.2	0.52*
mDBP	-0.60**	0.66**	0.65**	0.62**	0.56*	0.04	-0.33	0.21
stdDBT	-0.64**	0.66**	0.75***	0.67**	0.75***	-0.33	-0.24	0.52*
stdDBP	-0.57*	0.65**	0.61**	0.62**	0.56*	-0.02	-0.32	0.24
PH	0.36	-0.3	-0.38	-0.35	-0.52*	0.98***	0	-0.60**
PS	0.63**	-0.58*	-0.63**	-0.51*	-0.58*	0.15	0.74***	-0.23
TS	-0.53*	0.50*	0.53*	0.50*	0.56*	-0.68**	-0.57*	0.78***

Table 4.25. Correlation matrix of the breathing measures from the account creation and the survey task.

task, ($F(1, 16) = 17$, $p < .0008$, $R^2 = .515$) and ($F(1, 16) = 11.7$, $p < .0004$, $R^2 = .422$) respectively. The variability of troughs and peaks also correlates significantly between tasks r(16) = 67, p = .003, r(16) = 56, p = .017 and have mid to low predictive power ($F(1, 16) = 12.8$, $p < .003$, $R^2 = .444$) and ($F(1, 16) = 7.13$, $p < .017$, $R^2 = .308$). Finally the three last variables are, yet again highly correlated r(16) = .98, p = .00, r(16) = .74, p = .00 and r(16) = .78, p = .00 in the same order. The predictive power remains really high for peak height ($F(1, 16) = 465$, $p < .00$, $R^2 = .967$) but for peak ant trough sharpness it has dropped significantly when compared to the effect from the shopping task ($F(1, 16) = 19$, $p < .00$, $R^2 = .543$) and ($F(1, 16) = 24.4$, $p < .00$, $R^2 = .603$). It is clear that the relationship between the breathing patterns of these two tasks is close and on every measured variable suggesting a wide impact from the previous task. It would be interesting to see if this effect would grow stronger if the account creation process would be designed to induce more emotion, that way it would be possible to investigate the carryover effects to a greater extent. The effects from account creation is however not as strong as they have been from other tasks, notably the shopping task. The breathing patterns during the shopping task predicted respiration rate, peak height, peak sharpness and trough sharpness to a greater degree both for the account creation and for the survey tasks.

[17] * if the p-value is ≤ 0.05 (significant at the 5% level),
** if the p-value is ≤ 0.01 (significant at the 1% level),
*** if the p-value is ≤ 0.001 (significant at the 0.1% level).

S	RR	mDBT	mDBP	stdDBT	stdDBP	PH	PS	TS
Ac								
RR	0.549***[18]	0.557***	0.642***	0.536***	0.616***	0.13	0.10	0.22
mDBT	0.458**	0.515***	0.568***	0.538***	0.567***	0.15	0.04	0.275*
mDBP	0.361**	0.439**	0.422**	0.386**	0.314*	0.00	0.11	0.05
stdDBT	0.408**	0.431**	0.569***	0.444**	0.568***	0.11	0.06	0.269*
stdDBP	0.330*	0.422**	0.370**	0.385**	0.308*	0.00	0.10	0.06
PH	0.13	0.09	0.15	0.13	0.268*	0.967***	0.00	0.358**
PS	0.393**	0.336*	0.401**	0.259*	0.335*	0.02	0.543***	0.05
TS	0.278*	0.250*	0.278*	0.249*	0.316*	0.463**	0.320*	0.603***

Table 4.26. A matrix of the results from the (OLS) regression analysis for the account creation and survey task in the order of respiration rate (RR), mean distance between troughs (mDPT) and peaks (mDPT), the standard deviation between troughs (stdDBT) and peaks (stdDBP), the peak height (PH) and the sharpness of peaks (PS) and troughs (TS).

Summary. The regression analysis focused on understanding the influence of breathing patterns in one task on subsequent tasks, particularly in the context of emotional carryover effects. Six task pairs were analyzed: Character Creation and Shopping, Character Creation and Account Creation, Character Creation and Survey, Shopping and Account Creation, Shopping and Survey, and Account Creation and Survey. Each pair was examined for correlations and predictive power regarding respiration rates and patterns.

Character Creation and Shopping: Moderate correlation (r=0.48) and a significant, albeit limited, influence of Character Creation on Shopping's respiration rate (R^2=0.23). High correlations and predictive power were observed in peak height, peak sharpness, and trough sharpness between tasks.

Character Creation and Account Creation: Stronger correlation (r=0.57) compared to the first pair, indicating more similarity in breathing patterns. The influence on respiration rate was significant (R^2=0.33), and similarly high correlations in peak and trough measurements were noted.

Character Creation and Survey: No significant correlation in breathing patterns except for peak height and peak and trough sharpness measurements, which maintained high correlation and predictive power. The distancing of tasks in time likely contributed to the reduced predictive power for respiration rate.

Shopping and Account Creation: High correlation in respiration rates (r=0.83) and a strong predictive relationship (R^2=0.687) were observed, possibly due to the seamless transition between tasks. Peak and trough measurements also showed high correlation and predictive power.

Shopping and Survey: Medium to high correlation in breathing patterns and significant predictive power for respiration rate (R^2=0.612), with diminishing effects as tasks were further apart in time.

Account Creation and Survey: Significant correlations across all variables, indicating a close relationship in breathing patterns. Respiration rate showed a strong correlation (r=0.74) and predictive power (R^2=0.549). The emotional weight of the Account Creation task suggested potential carryover effects. Overall, varying degrees of influence of breathing patterns from one task to the next were found. The nature of tasks, their emotional weight, and the temporal proximity between tasks played significant roles in determining these relationships.

[18] ** if the p-value is ≤ 0.05 (significant at the 5% level),*
*** if the p-value is ≤ 0.01 (significant at the 1% level),*
**** if the p-value is ≤ 0.001 (significant at the 0.1% level).*

4.6. Qualitative responses and notes

In the discussion of qualitative feedback from the study, participants provided diverse insights and perspectives. That feedback was gathered from the optional open ended qualitative question at the end of the study and by informal discussions after the experiment where interesting insights were noted down. 6 out of 18 participants left a qualitative feedback at the end of the survey. They can be categorized into: problems with the study, improvements of the website and positive feedback. The most important of the three is problems with the study and those insights would impact a future iteration of the study. One participant was *"I was confused by the rating of the account creations. I was not sure what was meant. But I think I got it...".* This is referring to the SAM self-assessment scale that uses images instead of words. The reason this scale was used is that it has been used and validated over a long period of time and was meant to test the "test case" of the experiment as it was deployed right after the task was completed. It is clear now that it would have been better to use the same form of question for all the tasks. The images of the SAM scale were not clear enough and this feedback is helpful for pointing that out. This matter will be discussed more thoroughly in the limitations section. Another valuable feedback that happened early on in the experimentation phase was *"I could not put ''at'' on email address because the keyboard is different than the one I usually use, so I did not know where it is".* The participant in question could not find the @ sign on the keyboard as it had a different language setting than they were used to. This seems to have caused some anxiety because the participant in question rated their anxiety level as moderate to high on the questions. Their removal of the study was considered but since correlating their emotionality with the breathing patterns is still viable the data was kept. This feedback changed the procedure of the study for the remainder of the phase where participants were taught how to do the @ sign and made sure that the keyboard had the appropriate language setting.

A participant savvy in website development had useful pointers about the account creation task (the supposedly annoying one) and had his doubts that it would be annoying enough. They mentioned that frustration did not have a long enough time to build up and more problems should have been added before the error message about the password not having the correct symbols. They also mentioned that frustration usually comes when perception and results do not match, meaning that the Captchas used in the experiment were so complex and difficult that users would not expect to get them in the first attempt anyway. They would have been more frustrating if they were more clear and still would not work. Finally, they pointed out that clearing the form completely and refreshing the page when the error message showed up was such a common mistake in website development that people could be so used to it at this point that it does not cause significant frustration. These are all good tips for a future iteration of this study.

The insights about improvements to the website were more straightforward, one participant noted that it should *"specify what letters and symbols need to be in the password"* which is a good advice for website development but in the case of this experiment this was omitted purposefully to try to induce annoyance. Another one pointed out that *"Item that I was looking for in the store was not there, so I had to choose something else",* improving the amount of items in the store would give the experience more life and avoiding the problem of having people decide on an item for their persona beforehand and not finding it. This problem is very specific to this exact study but should be noted. Finally there is positive feedback with *"quite liked the study. I hope you have a lot from the data"* and "I genuinely thought that I was making mistakes during the captcha part. I got slightly annoyed but I thought it was my fault" which points out that the website was convincing. Informal discussions after the experiment suggested that some people enjoyed the experiment and many of them started talking about what clothes they found and "bought" and comparing it between other participants of the study. It appeared that the shopping experience, and browsing for clothes felt interesting and enjoyable. Some people were excited about the characters they created and a few of them went a great length to create a convincing character with a backstory that they wanted to discuss.

5. Discussion & Conclusion

This study investigated the relationship between emotional responses and breathing patterns during different web-based tasks, aiming to understand how emotions during one task influence subsequent tasks. A sample of 18 wore respiration belts, completed tasks, and provided emotional self-assessments. Results indicated potential links between breathing patterns and emotional responses, with certain tasks, particularly account creation and survey, showing significant emotional weight. The data analysis included survey response analysis, breathing pattern examination across tasks, and regression analysis to explore correlations and predictability for survey responses and breathing data between tasks. Although the sample size limited the findings, the study unveiled trends and relationships, suggesting the value of a larger-scale study. The qualitative feedback, categorized and enriched by expert consultation, underscored the need for study design adjustments and pointed toward future research directions.

The following section will go through the research question and the following sub-questions one by one, detailing the methods used to approach them and pointing out the primary findings. The findings will be linked to the literature and how they can answer the questions. As the combined answer to the four sub-questions answers the primary one there will be a summary of how they do in the end. Limitations of the study will be listed and discussed in detail followed by future research based on the findings and the experiment design. A conclusion subsection discussed the implications of the findings for the field suggesting practical applications.

The study's primary goal was to answer the research question "To what extent does the emotional response evoked by a task influence the emotional response to subsequent tasks, as measured by self-reported emotions and breathing patterns?" and sub-questions. Moreover, the study aims to explore the connection between breathing patterns and personal emotional evaluations. Using respiratory data to understand emotions is an intriguing objective, especially as it has not been a primary focus in Human-Computer interaction lately. Having participants go through well-known, and rather mundane, interfaces while tracking their breathing and emotions can give insights into how people respond to various interface features, how they respond when they do not work as intended, and how that affects their attitude and emotions in the following steps of their user journey. Even if the sample size is small it is relatively large compared to similar studies using breathing belts. The study is conducted as a rather small pilot study to investigate whether a research design like this can help answer these questions and aims to find promising measurements for respiration that can be helpful to achieve this goal. The results suggest that it could be interesting to conduct a similar study with a larger sample size and several adjustments.

Beginning to answer the first sub-question SQ1: In what ways do emotional responses, as measured by self-reported emotions and breathing patterns, vary across simple web-based tasks?" the analysis of the data gathered from the self-assessment scales revealed that there was a significantly higher (negative) emotional response, both for valence and arousal for the account creation task. This means that people reported being more anxious and annoyed during that task than the other two. This was to be expected since the task was designed with that in mind. Participants repeatedly attempted impossible CAPTCHA tasks after writing their account information and were then met by an error message suggesting that the problem was their fault, although it was out of their control. They then had to write everything again. The results indicate that this worked as intended to some extent. There was no significant difference between the other tasks suggesting that people did not differ in their emotions between the character creation task and after the survey.

To answer the second part of the question several respiration measurements were compared between the tasks. Respiration rate (RR) was found to be significantly higher while shopping when compared to the study as a whole as measured both by RR and the average differences between peaks and troughs. Unfortunately, there was no self-assessment question about the shopping experience by itself so there is no quantitative evidence for people rating their emotions differently during that task. There is, however, qualitative evidence of participants communicating after the study that they were excited about the clothes they found and wanted to discuss the shopping experience, while the other tasks did not come up in the same way. The shopping experience was described as ”enjoyable”. Changes in respiration rate have been shown to indicate a change in emotional state, with higher RR associated with emotional arousal whether it is excitement due to happiness, fear, or anger. Higher and more variable respiration rates have also been connected with amusement (Kreibig, 2010; Siddiqui et al., 2021). A formal self-evaluation of emotion by a larger sample, preferably happening in concurrence with, or right after the shopping experience could shed more light on these findings and offer explanations.

The sharpness of peaks and troughs differed from the whole experiment, both for the character creation task, the first one, and the survey, the last one. Differences in the rapidity of inhalation and exhalation and the ratio between them have been used as a marker for emotionality in the past. Quicker inhalations and slower exhalations are associated with anxiety (Kreibig, 2010). The results show that the sharpness of peaks and troughs is significantly lower during the survey, suggesting slower inhales and exhales respectively when compared to the study as a whole. The opposite was true for the sharpness of peaks during the character creation task where troughs were sharper, suggesting faster inhales, while the results for peaks were not significant. This could suggest that there was some level of heightened anxiety at the beginning of the experiment, reading the introduction and experiencing uncertainty while the realization that the experiment was over was considered a relief. Relief is connected with lower RR and slower exhales and inhales (Kreibig, 2010). These results are not significantly reflected in the self-assessment questions but show some promise for further investigation. It should be noted that self-reporting emotions has various limitations as noted in the literature review. It can be affected by memory limitations, various biases such as desirability bias, and simple unawareness of one’s emotional state (Pekrun, 2016; Robinson & Clore, 2002). Those limitations of self-assessment of emotions are important to consider here, although they are unlikely to be solely to blame for the discrepancies between the breathing data and the question data.

The study serves to answer the first sub-question. Participants experienced heightened negative emotions, particularly anxiety and annoyance, during the account creation task, aligning with the hypothesis following its design. In contrast, the shopping task seemed to induce positive engagement, suggested by a higher respiration rate and qualitative feedback, though without direct self-assessment data for confirmation. The character creation and survey tasks showed unique breathing patterns, with the former indicating potential initial anxiety and the latter suggesting a calming effect towards the experiment's end. These findings suggest some variation in emotional responses across different web-based tasks, as reflected in self-reported emotions and measured breathing patterns.

Depending on the different tasks, the study shows significant correlations between self-reported emotionality and some breathing measures, and various measurements of breathing significantly mirror the answers to the self-assessment of emotion questions. Similarly, there are some notable patterns in the responses to the questions that are also evidenced in the breathing data. There seem to be differences in how the questions asking about valence and arousal relate to the data which might indicate some interesting differences in the constructs themself.

The most interesting findings came from the responses to the last question asking for valence, question five. It showed significant mid-level correlations with respiration rates, distances between troughs and peaks, and the variability thereof. These variables were also predictive of a portion of the variance in the answers to the question to a significant extent, with RR explaining a third of it and the median variability explaining more than a fifth. This matches the literature nicely, but a large review published in 2010 explored physiological markers, including breathing, of various emotions based on multiple studies. The review names a higher respiration rate, and higher variations in inhalations and exhalations as critical markers for amusement, which is the closest emotion they have to "pleasant". "Joy" was the other candidate and it shares the increased respiration rate marker although there was no information on the variability. They name various other measures as well that would be highly interesting to test using this question in the future (Kreibig, 2010). The response from question four about the character creation task supports this finding where less pleased participants had longer pauses between their peak inhales suggesting lower respiration rate. The same goes for the responses to question two, asking about the valence during the account creation process, it showed that the peak sharpness decreased with annoyance which (as an estimate of exhalation time increase) fits with the literature where a lower ratio of inhalation time and a full breathing cycle is a marker of amusement.

The breathing data did not correlate as well with the responses to questions asking for arousal. The data was only predictive of the variance in responses to question three asking about arousal during the character creation task. Interestingly, this is the question that has the longest time delay from the task it is asking about. The level of arousal correlated significantly with the distance between inhales and the variability thereof, there was a notable correlation with peak sharpness as well. These results contradict what was hypothesized and what the literature on respiration related to anxiety tends to say where respiration rate tends to increase with anxiety. There is however evidence of increased variability in tidal volume which might be indicated by the variations in distance between troughs and peaks. The contrast with what the literature suggests raises questions about the time delta from the task to the question and if that could disturb the memory of the emotion felt at the time in some way.

The question asking about the valence at the end of the survey was the only question among the six that had notable correlations with the data from the experiment as a whole except for peak sharpness for most of the questions, although not to a $p = 0.05$ level significance. Various measures of the overall data could predict the variation in responses to the question to a significant extent. This question has the unique feature of asking about the participant's current state of valence at the end of the experiment. The relationship between self-reported valence and the respiration measures grows as the experiment runs and is at its strongest when asking about only the fourth task, while it is a little weaker for the data as a whole. This can have various implications. Valence seems to be related to current physiological experience, at least respiration. It could be difficult to accurately remember previous states of respiration and valence hence the most accurate response to the last question about the current state. None of the question responses hold any predictive power over the responses to the others for valence, mostly around 0.0, and there is no significant correlation, further suggesting that recalling valence could be difficult. The physiological traits related to valence could be cumulating over the course of the experiment and only at the end are they effective enough to be accurately represented in the last question and there it is representative to a lesser extent the overall experience. This would be highly interesting to study with multimodal measures of physical responses.

The relationship between self-assessed arousal and respiration does not seem to be as strong when it is inquired at the current moment and the strongest relationship was found when asking people to recall previous arousal levels. This fits interestingly with the correlations between the responses to the questions where the responses to the last question are significantly correlated with the other questions and there is a high correlation between the valence and arousal responses for the last questions. There seems to be internal consistency in how people respond to the questions about arousal with only the questions about the first task being correlated with the breathing data. That correlation did not match expectations or the literature. This could mean that participants are building their responses on an internal narrative or memory about their arousal state that is based on the beginning of the experiment and as they progress through the experiment this narrative influences their responses more than their physical state. An interesting addition to this would be if participants tended to misremember their arousal state from the first task and base their narrative on that. There is some evidence of this and would be interesting to test this theory in the future.

Finally, there is a relatively high correlation between the arousal and valence question pairs for the two tasks that were not rated as highly emotional while there is no correlation at all between the arousal and valence questions for the account creation task. This fits well with the literature and the circumplex model of affect or the valence-arousal model. It is a dimensional model that quantifies emotions as points on a two-dimensional plane with the center representing a neutral state. The dimension usually represented on the horizontal axis is valence, ranging from negative to positive, and arousal on the vertical ranging from high to low(Basu et al., 2015; Mauss & Robinson, 2009; Nandy et al., 2023). It is therefore expected that more extreme emotions are further away from each other on these axes. More neutral ones are closer to the center and closer to each other and this is reflected in the correlations of the answers. A similar, but much milder (and not significant) effect was observed in the correlations between self-reports and respiration where the differences were greatest between the questions related to account creation. This would be light evidence in favor of correspondence between self-reported emotions and breathing patterns.

Now there is evidence to answer the second sub-question: How does the pattern of breathing correspond to self-reported emotions across different web-based tasks? The results suggest that self-reported emotions have a significant and growing relationship with specific breathing patterns as participants progress through web-based tasks. However, it emphasized the complexity of the relationship. Respiration patterns were notably better at predicting responses to valence questions than arousal. Respiration metrics and valence responses align with literature, suggesting that breathing patterns reflect emotional states but the weaker correlation between arousal and breathing contradicts expectations and suggests an effect of memory or a narrative that influences the responses instead of respiration.

Addressing the last two sub-questions. These questions are aimed at understanding the relationships between the tasks in the context of the emotional markers and how they are reflected in the self-assessed emotions and breathing patterns across them. The sub-question: How predictive are initial task-induced responses, in terms of breathing patterns, and emotional self-assessments for subsequent task responses? This can be approached by looking at the findings from the regression analysis that show a complex relationship between the tasks. Looking at the self-assessment results there seems to be no predictive relationship between the responses to the questions regarding valence, this contrasts what was explained in the previous section that the overall breathing metrics were quite good at predicting the responses to the last valence question. Arousal on the other hand has much stronger predictive relationships between tasks where the arousal reported during the account creation task accounts for more than half the variance to the last question, and the character creation accounts for a fourth. This could imply a recency effect where the tasks that are closer to each other in time have a more similar arousal response. Character creation, however, has a much lower predictive relationship with the third task, account creation than the fourth which does not support that hypothesis. The highest predictive relationship is between account creation, the task that had the highest emotional weight, and the last question where people were asked about their current arousal after a neutral survey. This hints at an emotional carryover effect on top of the recency effect where the emotional strength of the task impacts how much of the emotion is present in the following task.

Looking at the same relationships in the context of breathing data. Similar effects appear in the data, tasks that are closer together have higher correlations and stronger predictive relationships. For example, the shopping task and account creation are adjacent and are highly correlated on multiple metrics with strong predictive relationships, most notably for respiration rate. This relationship can be simply due to the recency effect. The shopping task had a significantly higher respiration rate than the other tasks and there was qualitative evidence that it was enjoyable, although the question was lacking from the study, this could therefore be an example of a carryover of a strong emotion but would require more evidence. In contrast, there are hardly any correlations between the first task (character creation) and the last, survey. Shopping and survey have various high correlations but they are all lower than the ones shopping had with account creation suggesting a fading effect. account creation had a strong correlation and a predictive relationship on all metrics breathing strongly implying that the most emotional task had a significant effect on the end of the experiment.

Specific breathing measures, notably peak height (PH), peak sharpness (PS), and trough sharpness (TS), demonstrate a consistently high correlation and predictive power across all tasks. This could imply that these are robust indicators of physiological and emotional state and that they can predict those in the following tasks. This could also simply mean these are fairly stable overall and do not change much depending on these tasks, but they all have variability in the overall data. There is however a notable variability in their predictive power depending on the tasks. The relationships are consistently stronger the closer the tasks are in time, so character creation has a much stronger correlation and predictive relationship for these three variables for shopping, the second task, than it does for the survey, the fourth one, and so on. This might then mean that these measures show high predictability while others have more variability in their predictive power. This suggests that some aspects of breathing patterns are more consistently linked to emotional states than others where predictiveness can fluctuate more depending on the task and the distance between them. It will be interesting to look for different variables that can become even better predictors for breathing in other tasks.

To sum up the answer to the third sub-question, initial task-induced responses, in terms of breathing patterns and emotional self-assessments, exhibit varied predictive power for subsequent task responses. Self-assessed valence shows limited predictive relationships between tasks while arousal has stronger predictive links which are heightened when the tasks are close in time and when the emotionality of them is strong. Breathing data analysis revealed similar patterns. Tasks that are temporarily closer have higher correlations and higher predictive relationships and the same is true when the former task elicits is rated highly emotional by the self-assessment measure. There was a clear difference between the predictive power of different metrics with some showing a robust predictability between tasks while others fluctuated depending on other features.

The same insights can be used to answer the fourth sub-question: What are the implications of the observed relationships for understanding emotional carry-over effects in web-based tasks? The findings suggest that emotional carry-over effects were present that can be detected through self-reported emotions and breathing analysis. The varying degrees of predictive power and correlations between tasks emphasize the effects emotional intensity, task characteristics and the temporal distance between tasks have on the emotional experience of participants. Some breathing measures, PH, PS, and TS constantly had high predictive power between tasks and could have potential as relative indicators of emotional state. Others, such as RR, were not as robust but displayed the same effects of carrying over to the next tasks in the same way.

The culmination of the answer to the four sub-questions serves to answer the primary one: "To what extent does the emotional response evoked by a task influence the emotional response to subsequent tasks, as measured by self-reported emotions and breathing patterns?". As has been stated in detail above emotional response evoked by a task does seem to influence the response to subsequent tasks to a significant degree. It is however dependent on the nature of the task, and the time between the tasks. It is also dependent on the emotions in question, emotions more related to valence than arousal do not tend to significantly carry over between tasks to the same extent as those related to arousal according to the self-evaluations of emotions. Some breathing patterns do seem to change between tasks more than others and do seem to follow trends in markers related to certain emotions according to the literature.

5.1. Limitations & future research

This study, while comprehensive in many aspects, is not without its limitations. Throughout this study design things were noted that could be different and that can advise future iterations of the study. These limitations stem primarily from a small sample size, technological limitations, and some decisions made in the study design. Each of these limitations, discussed in detail below, provides context for the interpretation of the findings and suggests directions for future research. The number of participants (N= 18) is small to get significant results for the study. However, it was to show various relationships and trends in the data and show that the experiment design has some value. Its limitations became apparent when trying to compare the results of different answers to the emotional self-assessment questions where a significant part of the answers had only one, or even no answer rendering many statistical rests unusable and less accurate. This would be more unlikely to happen with a bigger sample size and could broaden the scope of the results with a more varied response to the questions. With the low number of participants, the data usually showed high variability, a larger sample size could lower the standard deviations and make the results more generalizable.

Certain limitations became apparent during, and after the data-gathering phase. Although instructed not to a relatively big portion of the participants accidentally spoke out during the experiment. This included silent mumbling to themselves and questions toward the researcher. They usually realized that they were not supposed to speak. It appears that some people tend to quietly talk to themselves when doing tasks like the ones in the experiment. Although these were few and far between they would be registered as a change in breathing. The same can be said about coughing, which also happened with several participants throughout the experiment. These incidents were noted down by the researcher simply by stating what happened and during what part of the experiment. When it then came to using that information, looking for these anomalies in the data it turned out it was not usable due to an insufficient temporal accuracy, so the removal of the removal of the cough or the speech from the data was not possible. While these incidents do not have a great impact on the average measurements of the data, since they tend to be less than a second long in 15 minutes of data, they are a limitation that can be improved. Recording audio and preferably video of the participant during the experiment could solve this problem. A video with an accurate timestamp could give the exact moment an incident happens and when it ends and would allow for a full removal of it from the dataset.

Another limitation concerning the data gathering was only discovered after all the data had been gathered and when it was to be normalized. The normalization was based on the calibration period, at the beginning of the experiment, when participants were instructed to "box breathe", and slowly take deep breaths with a three-second delay. When looking at the data there were a few datasets that showed the tendency to report high values at the beginning of the experiment and after 50-70 seconds drop significantly for the rest of the experiment. This made it difficult to use the data from the calibration period as most of it was higher than the rest of the data. It was however for most of them to identify peaks and troughs sufficient for the normalization, except one participant that had to be removed from the data. The implications of this could be that some of the datasets are missing some of the lowest troughs as they cap out at 0. Care was taken to keep that in moderation and mainly to those that would likely be flagged as outliers. This phenomenon was inspected thoroughly and it tends to occur when the participant has the breathing belt on rather tightly and begins the experiment sitting upright. If the participant then hunches down towards the screen the belt changes slightly, but enough to cause this drop. In future iterations, it should be firmly added to the instructions to sit upright throughout the experiment.

The last limitations to be mentioned that have to do with the data gathering concerns the questionnaire at the end of the experiment. It was decided to use the SAM scale for the first two questions, the reason for that choice was that the SAM scale is a tried and tested scale to assess emotions concurrently or close to it. The scale and the images are quite old and like qualitative feedback received from a participant, it was not completely clear. The problem with the SAM scale being dated has been brought up in the literature before (Betella & Verschure, 2016). The following questions are on a 7-point Likert scale using words to signal the level of emotion while SAM is on a 9-point scale and uses images. This discrepancy between scales risks adding unnecessarily to the variance of the answers and should be avoided. A better practice in a future study would be to stick to the same form of questions for all questions, and judging by this study, prefer the worded ones. Finally, the scale had 7 questions when it would have been advantageous to have 9. Questions 3 and 4 ask about the character creation and were supposed to cover the character creation and the shopping experience. This however turned out to be such a long task that it was over twice the length of the others. It was therefore decided to split it into two tasks, now all of them were of a relatively equal length, and the breathing patterns of the shopping task could be inspected on their own. They could however not be compared to the results from the self-assessment scale except for in the context of the character creation questions. Adding specific questions for the shopping task could greatly add to the insights of the experiment.

The website itself had some potential limitations worth noting. As pointed out by a UI designer in the qualitative data, the design of the account creation might not be sufficient to induce annoyance or anxiety in participants with experience using these types of forms. That applies to most of the participants. The task should have been more focused on making the user believe they were doing everything correctly and then breaking that trust, the CAPTCHAS should have looked simple and still not work. In future iterations of the study, it would be interesting to see an increase in the emotionality of the high-emotion task, preferably by conducting detailed user studies beforehand to maximize for a certain emotion. That way the contrast could be more clear and the results hopefully be more pronounced and for more types of breathing patterns.

The study was designed with that in mind that the website experience should be authentic and believable. That meant designing an online store that was plausible and with various inner sites and information texts, a location, and a history. This was done, both to get an idea of participants' state and respiration in a believable real-world-like environment and to remove suspicion of any foul play during the account creation where it was desirable that the participant did not realize they were being manipulated with the error messages. This however has its downside that can be considered a limitation, the experience of the participants is not the same. There was a high variation in how much participants explored the website, they all had the same task to go to the shop and buy an item, and every site had multiple links to the shop but some people liked exploring. This meant that the time between the first task and the rest of them was not the same for all participants and the shopping task was not the same for them all. Some spent their time reading and looking at maps while others looked at multiple pieces of clothing. While this was in part intentional it could be helpful for a future study to design the tasks so they are more comparable between people and more concrete.

Finally, two participants consistently rated themselves as slightly or mostly anxious throughout the experiment. It is possible that their anxiety had to do with something outside of the experiment, like an anxiety disorder. The effects of those could have been mediated using anxiety disorder assessment tests. It is also a possibility that the participants were simply shy or uncomfortable doing the experiment for some unrelated reason that could cause heightened levels of anxiety. Some people had to have the breathing belt adjusted more than others and that might have caused stress, there are various possibilities.

5.1.1. Future research

The results indicate that there is a potential for breathing patterns to be used to estimate or even predict responses on emotional self-assessment scales. With greater advancement of this knowledge, it could be used to replace the scales altogether. With physiological estimates of emotions becoming more prevalent accurate breathing patterns will add a valuable dimension. Making this possible would require studies on a much bigger scale with very large sample sizes. A much bigger study could show more robust correlations between the self-reporting of emotion and the various respiration measurements. A study like that could include more measurement variables that were, in some cases, ignored for this study to avoid it becoming too big due to time constraints. Examples of measurements that would be interesting to measure are relative volume of breath, this can be done by calculating the area under the curve from peak to trough. This could give an estimate of how much air enters the lounge in a given breath. Other examples could be more precise measures of variations of breath, where a given task could be split up into even finer subtasks, or even simply by calculating the variability for more variables. Measuring a breath symmetry index, by comparing either the duration or the area under the curve for inhales and exhales and finding the ratio. This measurement could give some ideas about emotionality as it is a commonly applied strategy in yoga and with breathing practices to control this ratio. More measures could include entropy measures and the ratio between inhalation and exhalation duration. The possibilities are vast and with a big study it would be interesting to uncover them.

Future studies could build on a similar study design but with different tasks. The tasks could be designed to evoke different emotions and in different order to further study the potential carryover effects. A large-scale between-subjects design for the study could give better insight into the correlations between the respiration patterns connected with different emotionality. Being able to see how the patterns related to annoyance correlate with the ones related to anxiety would be an improvement in this study that would require a substantially larger dataset. A between-subjects experiment would similarly be better suited to determine the extent of the carryover effects by having a control group that gets an emotionally neutral task instead of a highly emotional one.

Studies that would focus more on finding the relationships between various breathing patterns and the results from the self-assessment questions could use the results from this one that showed a much greater relationship when the question was answered concurrently with the emotion compared to shortly or moderately long after. In a study that does not focus so much on task switching or carryover effect every task could include an emotional self-assessment question.
It could be valuable to add other physiological measures of emotion to the design for an increased number of points on the nomological network for emotion. Measuring heart rate and skin conductance, for example, alongside respiration and comparing the figures with a self-assessment scale could give more robust results. Correlating each of the variations with each other and looking for concurrent trends would be informative and could move the research in the direction of being fully focused on physiological measurements.

A more futuristic, but not unreasonable, future research would apply a similar study design but exclude the breathing belt. There has already some research been done on the detection of respiration using a non-invasive radio ultra-wideband radar using machine learning (Siddiqui et al., 2021). If respiration patterns could be detectable via a microphone this could raise the importance of this research to a new level where a non-invasive method of breath detection could be a part of a normal laptop allowing for emotion estimation while a user uses a website or an application giving data that can be used to streamline the experience based on their mood or feelings. Using information on carryover effects and breathing patterns it would be possible to rearrange certain tasks to minimize the effects of previous emotionally heightened tasks using some sort of real-time prediction and feedback system. Machine learning algorithms could help improve this field of study significantly with pattern recognition. Machine learning could greatly improve the nuances, complexities, and non-linear properties of the breathing patterns possible to detect and correlate leading to more accurate emotion prediction models.

There is much potential for and value in research of breathing patterns concerning emotion. This relatively underexplored modality can be a great asset to the multifaceted arsenal of human-computer interaction research. With the progression of AI models, it will only become more powerful and relevant and it will be exciting to watch how breathing research will develop within the field of HCI.

5.2. Conclusion

The study's findings underscore the potential for breath analysis in emotional research in the field of HCI and in other fields. With its limited sample size, it managed to show clear evidence of a relationship between different breathing markers, such as respiration rate or sharpness of breath, and self-reported emotional data. Moreover, the results showed a difference in emotional ratings between different web-based tasks and how that was consistent with respiration data for several markers. There was evidence that the emotional effect of an initial task impacts people's emotionality during the subsequent tasks as noted both by self-assessment and breathing changes. The study can be considered a pilot study where the value of various known breathing patterns was tested in terms of their usability as markers for emotion, while many were left out. The research design was also being tested and the feasibility of its use in evoking emotion to the level of being measurable, and while it was successful in some ways others can be greatly improved.

The results emphasize the possibilities for HCI research into using breathing patterns as markers for emotion. The simple data from a breathing belt can be analyzed in a vast amount of ways to abstract different elements of breathing each of them potentially a useful marker for emotionality. Multiple studies of a similar nature across the field could investigate a large number of emotions and their relationship with breathing to create a library of this information. That opens up a lot of opportunities where, for example, non-human agents can breathe correctly so it fits their displayed emotions. A library like that could also be the other way around and identify people's emotions based on their respiration. The implications of this study could prove useful for user experience design and studies. Developers and designers might integrate breathing pattern analysis into their systems, enabling real-time emotional tracking without the intrusive nature of self-assessment methods. This could lead to more adaptive and responsive interfaces that could adjust the content, the order of it, or its difficulty based on the user's current emotional state, enhancing user satisfaction and engagement. Good estimations of the emotional impacts of different tasks could incentivize developers to place tasks that are deemed emotionally sensitive further away from emotionally laden tasks based on the results gathered in the user studies. A frustrating account creation process should maybe not be placed right before a shopping experience or a checkout where frustration could influence people's behavior.

The use of biometric data to track people's emotions requires robust policies around it and these results and future breathing research could help pave the way for new standards in emotional data collection and analysis. Ethical concerns regarding privacy and consent must have clear guidelines and a balance must be struck between technological advancements and individual rights. If this technology is treated with care it could have a broad set of applications in the field of HCI.

The objectives of this study were met in large part. The hypothesis that differences in how people self-reported their emotional states could, at least in part, be seen in their breathing data, and the results show evidence of that. The clear difference in self-reported emotion relating to the account creation task, designed to invoke negative emotion was in support of that hypothesis. The same applies to how emotional state carries over to the subsequent tasks, how it is more similar the closer they are in time and the stronger the emotions are. The connection between various respiration markers and self-reported emotion was in line with the literature suggesting the measurements have some grounding. There is very little literature on emotional tracking via respiration in the HCI literature and this study could pave the way for more research in this area, viewing user and interface studies from a new perspective.

Standing on the intersection of technology and human emotion this study offers guidance towards a future where digital interfaces do not simply respond to clicks or swipes but to the rhythm of our breath, adapting to our emotional needs. This research emphasizes the power of interdisciplinary research combining the fields of computer science and psychology to further our understanding of the complexities of human emotion. This study should be viewed as a reminder of the potential for new technology that does not just understand our commands but also our feelings, making sure that future digital interactions are not only efficient but emotionally intelligent, considerate, and human.

6. Bibliography

A, R. Banu., Al Siyabi, W. S. A., & Al Minje, Y. (2021). A Conceptual Review on Integration of Cognitive Load Theory and Human-Computer Interaction. *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, 667–672. https://doi.org/10.1109/ICSECS52883.2021.00127

Abbruzzese, L., Magnani, N., Robertson, I. H., & Mancuso, M. (2019). Age and Gender Differences in Emotion Recognition. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02371

Adler, R. F., & Benbunan-Fich, R. (2013). Self-interruptions in discretionary multitasking. *Computers in Human Behavior*, *29*(4), 1441–1449. https://doi.org/10.1016/j.chb.2013.01.040

Al-Shawaf, L., Conroy-Beam, D., Asao, K., & Buss, D. M. (2016). Human Emotions: An Evolutionary Psychological Perspective. *Emotion Review*, *8*(2), 173–186. https://doi.org/10.1177/1754073914565518

Appelhans, B. M., & Luecken, L. J. (2006). Heart Rate Variability as an Index of Regulated Emotional Responding. *Review of General Psychology*, *10*(3), 229–240. https://doi.org/10.1037/1089-2680.10.3.229

Backs, R., da Silva, S., & Han, K. (2005). A Comparison of Younger and Older Adults' Self-Assessment Manikin Ratings of Affective Pictures. *Experimental Aging Research*, *31*, 421–440. https://doi.org/10.1080/03610730500206808

Badawy, J., Nguyen, O. K., Clark, C., Halm, E. A., & Makam, A. N. (2017). Is everyone really breathing 20 times a minute? Assessing epidemiology and variation in recorded respiratory rate in hospitalised adults. *BMJ Quality & Safety*, *26*(10), 832–836. https://doi.org/10.1136/bmjqs-2017-006671

Balestrino, M., & Somjen, G. G. (1988). Concentration of carbon dioxide, interstitial pH and synaptic transmission in hippocampal formation of the rat. *The Journal of Physiology*, *396*, 247–266.

Basu, S., Jana, N., Bag, A., M, M., Mukherjee, J., Kumar, S., & Guha, R. (2015). Emotion recognition based on physiological signals using valence-arousal model. *2015 Third International Conference on Image Information Processing (ICIIP)*, 50–55. https://doi.org/10.1109/ICIIP.2015.7414739

Bernardet, U., Kang, S., Feng, A., DiPaola, S., & Shapiro, A. (2017). A Dynamic Speech Breathing System for Virtual Characters. In J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, & S. Kopp (Eds.), *Intelligent Virtual Agents* (pp. 43–52). Springer International Publishing. https://doi.org/10.1007/978-3-319-67401-8_5

Betella, A., & Verschure, P. F. M. J. (2016). The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLOS ONE*, *11*(2), e0148037. https://doi.org/10.1371/journal.pone.0148037

Boiten, F. A., Frijda, N. H., & Wientjes, C. J. E. (1994). Emotions and respiratory patterns: Review and critical analysis. *International Journal of Psychophysiology*, *17*(2), 103–128. https://doi.org/10.1016/0167-8760(94)90027-2

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brosch, T., Scherer, K., Grandjean, D., & Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Medical Weekly*, *143*(1920), Article 1920. https://doi.org/10.4414/smw.2013.13786

Bynion, T.-M., & Feldner, M. (2017). *Self-Assessment Manikin* (pp. 1–3). https://doi.org/10.1007/978-3-319-28099-8_77-1

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, *50*(4), 31–34. https://doi.org/10.1145/1232743.1232769

Chien, H.-C., Chung, Y.-C., Yeh, M.-L., & Lee, J.-F. (2015). Breathing exercise combined with cognitive behavioural intervention improves sleep quality and heart rate variability in major depression. *Journal of Clinical Nursing*, *24*(21–22), 3206–3214. https://doi.org/10.1111/jocn.12972

Coleman, A., & Snarey, J. (2011). James-Lange Theory. *Encyclopedia of Child Behavior and Development*, *2*, 844–846.

Cools, R. (2015). Neuropsychopharmacology of Cognitive Flexibility. In A. W. Toga (Ed.), *Brain Mapping* (pp. 349–353). Academic Press. https://doi.org/10.1016/B978-0-12-397025-1.00253-0

Critchley, H., & Nagai, Y. (2013). Electrodermal Activity (EDA). In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (pp. 666–669). Springer. https://doi.org/10.1007/978-1-4419-1005-9_13

David Bowman, N., Keene, J., & Najera, C. J. (2021). Flow Encourages Task Focus, but Frustration Drives Task Switching: How Reward and Effort Combine to Influence Player Engagement in a Simple Video Game. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–8. https://doi.org/10.1145/3411764.3445678

Del Negro, C. A., Funk, G. D., & Feldman, J. L. (2018). Breathing matters. *Nature Reviews. Neuroscience*, *19*(6), 351–367. https://doi.org/10.1038/s41583-018-0003-6

Demanet, J., Liefooghe, B., & Verbruggen, F. (2011). Valence, arousal, and cognitive control: A voluntary task-switching study. *FRONTIERS IN PSYCHOLOGY*, *2*, 1–9. https://doi.org/10.3389/fpsyg.2011.00336

Demangeot, C., & Broderick, A. J. (2007). Conceptualising consumer behaviour in online shopping environments. *International Journal of Retail & Distribution Management*, *35*(11), 878–894. https://doi.org/10.1108/09590550710828218

Demartini, Z., Maranha Gatto, L. A., Koppe, G. L., Francisco, A. N., & Guerios, E. E. (2020). Ondine's curse: Myth meets reality. *Sleep Medicine: X*, *2*, 100012. https://doi.org/10.1016/j.sleepx.2020.100012

Dreisbach, G., & Goschke, T. (2004). How Positive Affect Modulates Cognitive Control: Reduced Perseveration at the Cost of Increased Distractibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 343–353. https://doi.org/10.1037/0278-7393.30.2.343

Drummond, G. B., Fischer, D., & Arvind, D. K. (2020). Current clinical methods of measurement of respiratory rate give imprecise values. *ERJ Open Research*, *6*(3). https://doi.org/10.1183/23120541.00023-2020

Duda | The Professional Website Builder You Can Call Your Own. (n.d.). Retrieved 1 February 2024, from https://www.duda.co/

Eich, E., Geraerts, E., Schooler, J. W., & Forgas, J. P. (2008). 2.15—Memory in and about Affect. In J. H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 239–260). Academic Press. https://doi.org/10.1016/B978-012370509-9.00146-7

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, *99*(3), 550–553. https://doi.org/10.1037/0033-295x.99.3.550

Fredrickson, B. L. (2004). The broaden–and–build theory of positive emotions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1449), 1367–1377. https://doi.org/10.1098/rstb.2004.1512

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion*, *19*(3), 313–332. https://doi.org/10.1080/02699930441000238

Frogner, J. I., Noori, F. M., Halvorsen, P., Hicks, S. A., Garcia-Ceja, E., Torresen, J., & Riegler, M. A. (2019). One-Dimensional Convolutional Neural Networks on Motor Activity Measurements in Detection of Depression. *Proceedings of the 4th International Workshop on Multimedia for Personal Health & Health Care*, 9–15. https://doi.org/10.1145/3347444.3356238

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do Girls Really Experience More Anxiety in Mathematics? *Psychological Science*, *24*(10), 2079–2087. https://doi.org/10.1177/0956797613486989

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, *53*(2), 91–103. https://doi.org/10.1016/j.ijpsycho.2004.02.002

Hart, S. J., Green, S. R., Casp, M., & Belger, A. (2010). Emotional priming effects during Stroop task performance. *NeuroImage*, *49*(3), 2662–2670. https://doi.org/10.1016/j.neuroimage.2009.10.076

Homma, I., & Masaoka, Y. (2008). Breathing rhythms and emotions. *Experimental Physiology*, *93*(9), 1011–1021. https://doi.org/10.1113/expphysiol.2008.042424

Hsieh, S., & Lin, S. J. (2019). The Dissociable Effects of Induced Positive and Negative Moods on Cognitive Flexibility. *Scientific Reports*, *9*, 1126. https://doi.org/10.1038/s41598-018-37683-4

Jeon, M. (2017). Chapter 1 - Emotions and Affect in Human Factors and Human–Computer Interaction: Taxonomy, Theories, Approaches, and Methods. In M. Jeon (Ed.), *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 3–26). Academic Press. https://doi.org/10.1016/B978-0-12-801851-4.00001-X

Josef Perktold, Skipper Seabold, Kevin Sheppard, ChadFulton, Kerby Shedden, jbrockmendel, j-grana6, Peter Quackenbush, Vincent Arel-Bundock, Wes McKinney, Ian Langmore, Bart Baker, Ralf Gommers, yogabonito, s-scherrer, Yauhen Zhurko, Matthew Brett, Enrico Giampieri, yl565, … Yaroslav Halchenko. (2023). *statsmodels/statsmodels: Release 0.14.1* (v0.14.1) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.593847

Kaplan, S. (1975). *An informal model for the prediction of preference*. Dowden, Hutchinson and Ross. http://deepblue.lib.umich.edu/handle/2027.42/150732

Kensinger, E. A. (2009). Remembering the Details: Effects of Emotion. *Emotion Review*, *1*(2), 99–113. https://doi.org/10.1177/1754073908100432

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*, 849–874. https://doi.org/10.1037/a0019842

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 213–233. https://doi.org/10.1007/s11412-018-9277-y

Klausen, T. A., Farhadi, U., Vlachos, E., & Jorgensen, J. (2022). Signalling Emotions with a Breathing Soft Robot. *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*, 194–200. https://doi.org/10.1109/RoboSoft54090.2022.9762140

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kučera, D., & Haviger, J. (2012). Using Mood Induction Procedures in Psychological Research. *Procedia - Social and Behavioral Sciences*, *69*, 31–40. https://doi.org/10.1016/j.sbspro.2012.11.380

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K. J., & Russell, J. A. (2017). The Relation Between Valence and Arousal in Subjective Experience Varies With Personality and Culture: Relation of Valence to Arousal. *Journal of Personality*, *85*(4), 530–542. https://doi.org/10.1111/jopy.12258

Lee, K., Choi, J., Marakas, G. M., & Singh, S. N. (2019). Two distinct routes for inducing emotions in HCI design. *International Journal of Human-Computer Studies*, *124*, 67–80. https://doi.org/10.1016/j.ijhcs.2018.11.012

Lee, Y., & Kozar, K. (2009). Designing usable online stores: A landscape preference perspective. *Information & Management*, *46*, 31–41. https://doi.org/10.1016/j.im.2008.11.002

Levine, L. J., & Burgess, S. L. (1997). Beyond General Arousal: Effects of Specific Emotions on Memory. *Social Cognition*, *15*(3), 157–181. https://doi.org/10.1521/soco.1997.15.3.157

Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition*, *22*, 530–554. https://doi.org/10.1521/soco.22.5.530.50767

Lim, N. (2016). Cultural differences in emotion: Differences in emotional arousal level between the East and the West. *Integrative Medicine Research*, *5*(2), 105–109. https://doi.org/10.1016/j.imr.2016.03.004

Liu, H., Fan, N., Rossi, S., Yao, P., & Chen, B. (2016). The effect of cognitive flexibility on task switching and language switching. *International Journal of Bilingualism*, *20*(5), 563–579. https://doi.org/10.1177/1367006915572400

Magnusson, K. R., & Brim, B. L. (2014). The Aging Brain. In *Reference Module in Biomedical Sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-801238-3.00158-6

Masaoka, Y., Izumizaki, M., & Homma, I. (2014). Where is the rhythm generator for emotional breathing? *Progress in Brain Research*, *209*, 367–377. https://doi.org/10.1016/B978-0-444-63274-6.00019-9

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

McGaugh, J. L. (2000). Memory—A Century of Consolidation. *Science*, *287*(5451), 248–251. https://doi.org/10.1126/science.287.5451.248

Nabizadeh Chianeh, G., Vahedi, S., Rostami, M., & Nazari, M. A. (2012). *Validity and Reliability of Self-Assessment Manikin*. *6*(2), 52–61.

Nandy, R., Nandy, K., & Walters, S. T. (2023). Relationship Between Valence and Arousal for Subjective Experience in a Real-life Setting for Supportive Housing Residents: Results From an Ecological Momentary Assessment Study. *JMIR Formative Research*, *7*, e34989. https://doi.org/10.2196/34989

Novick, D., Afravi, M., & Camacho, A. (2018). *PaolaChat: A Virtual Agent with Naturalistic Breathing* (pp. 351–360). https://doi.org/10.1007/978-3-319-91581-4_26

Oh, H., Yun, Y., & Myung, R. (2021). Cognitive Modeling of Task Switching in Discretionary Multitasking Based on the ACT-R Cognitive Architecture. *Applied Sciences*, *11*(9), Article 9. https://doi.org/10.3390/app11093967

Pace-Schott, E. F., Amole, M. C., Aue, T., Balconi, M., Bylsma, L. M., Critchley, H., Demaree, H. A., Friedman, B. H., Gooding, A. E. K., Gosseries, O., Jovanovic, T., Kirby, L. A. J., Kozlowska, K., Laureys, S., Lowe, L., Magee, K., Marin, M.-F., Merner, A. R., Robinson, J. L., … VanElzakker, M. B. (2019). Physiological feelings. *Neuroscience & Biobehavioral Reviews*, *103*, 267–304. https://doi.org/10.1016/j.neubiorev.2019.05.002

pandas.DataFrame.rolling—Pandas 2.1.4 documentation. (n.d.). Retrieved 9 January 2024, from https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rolling.html

Paulitzki, J. R., Risko, E. F., Oakman, J. M., & Stolz, J. A. (2008). Doing the unpleasant: How the emotional nature of a threat-relevant task affects task-switching. *Personality and Individual Differences*, *45*(5), 350–355. https://doi.org/10.1016/j.paid.2008.05.003

Pekrun, R. (2016). Using Self-Report to Assess Emotions in Education. In M. Zembylas & P. A. Schutz (Eds.), *Methodological Advances in Research on Emotion and Education* (pp. 43–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-29049-2_4

Pessanha, F., Kaya, H., Akdag Salah, A. A., & Salah, A. A. (2022). Towards using Breathing Features for Multimodal Estimation of Depression Severity. *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*, 128–138. https://doi.org/10.1145/3536221.3556606

Pham, M. T. (2007). Emotion and Rationality: A Critical Review and Interpretation of Empirical Evidence. *Review of General Psychology*, *11*(2), 155–178. https://doi.org/10.1037/1089-2680.11.2.155

Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*(2), 198–202. https://doi.org/10.1016/j.conb.2004.03.015

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–734. https://doi.org/10.1017/S0954579405050340

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Roes, R. H., Pessanha, F., & Akdag Salah, A. (2022a). An Emotional Respiration Speech Dataset. *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*, 70–78. https://doi.org/10.1145/3536220.3558803

Roes, R. H., Pessanha, F., & Akdag Salah, A. (2022b). An Emotional Respiration Speech Dataset. *Companion Publication of the 2022 International Conference on Multimodal Interaction*, 70–78. https://doi.org/10.1145/3536220.3558803

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Roter, D. L., Frankel, R. M., Hall, J. A., & Sluyter, D. (2006). The Expression of Emotion Through Nonverbal Behavior in Medical Visits: Mechanisms and Outcomes. *Journal of General Internal Medicine*, *21*(Suppl 1), S28. https://doi.org/10.1111/j.1525-1497.2006.00306.x

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1819–1828. https://doi.org/10.1145/1518701.1518981

Sarkar, A. A. (2017). Functional Correlation between Breathing and Emotional States. *MOJ Anatomy & Physiology*, *3*(5). https://doi.org/10.15406/mojap.2017.03.00108

Sauter, D. A. (2017). The Nonverbal Communication of Positive Emotions: An Emotion Family Approach. *Emotion Review*, *9*(3), 222–234. https://doi.org/10.1177/1754073916667236

Schestatsky, P., & Fernandes, L. N. T. (2004). Acquired Ondine's curse: Case report. *Arquivos De Neuro-Psiquiatria*, *62*(2B), 523–527. https://doi.org/10.1590/s0004-282x2004000300027

Schmitz, F., & Voss, A. (2014). Components of task switching: A closer look at task switching and cue switching. *Acta Psychologica*, *151*, 184–196. https://doi.org/10.1016/j.actpsy.2014.06.009

scipy.signal.find_peaks—SciPy v1.11.4 Manual. (n.d.). Retrieved 23 November 2023, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

Siddiqui, H. U. R., Shahzad, H. F., Saleem, A. A., Khan Khakwani, A. B., Rustam, F., Lee, E., Ashraf, I., & Dudley, S. (2021). Respiration Based Non-Invasive Approach for Emotion Recognition Using Impulse Radio Ultra Wide Band Radar and Machine Learning. *Sensors (Basel, Switzerland)*, *21*(24), 8336. https://doi.org/10.3390/s21248336

Siedlecka, E., & Denson, T. F. (2019). Experimental Methods for Inducing Basic Emotions: A Qualitative Review. *Emotion Review*, *11*(1), 87–97. https://doi.org/10.1177/1754073917749016

Sternglanz, R. W., & DePaulo, B. M. (2004). READING NONVERBAL CUES TO EMOTIONS: THE ADVANTAGES AND LIABILITIES OF RELATIONSHIP CLOSENESS. *Journal of Nonverbal Behavior*, *28*(4), 245–266. https://doi.org/10.1007/s10919-004-4158-7

Talarico, J. M., Berntsen, D., & Rubin, D. C. (2009). Positive emotions enhance recall of peripheral details. *Cognition and Emotion*, *23*(2), 380–398. https://doi.org/10.1080/02699930801993999

The Cost of Not Paying Attention: How Interruptions Impact Knowledge Worker Productivity | Information Overload Research Group. (2010, April 5). https://iorgforum.org/service/the-cost-of-not-paying-attention-how-interruptions-impact-knowledge-worker-productivity/

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The Influences of Emotion on Learning and Memory. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01454

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Wang, Y., Chen, J., & Yue, Z. (2017). Positive Emotion Facilitates Cognitive Flexibility: An fMRI Study. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01832

Yang, X., Ram, N., Robinson, T., & Reeves, B. (2019). Using Screenshots to Predict Task Switching on Smartphones. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3313089

Zhou, Y., & Fischer, M. H. (2018). Mimicking non-verbal emotional expressions and empathy development in simulated consultations: An experimental feasibility study. *Patient Education and Counseling*, *101*(2), 304–309. https://doi.org/10.1016/j.pec.2017.08.016

Appendix 1 – Consent form

Welcome

Thank you for considering participation in my experiment. The experiment is being conducted as part of my master's thesis in the Human-Computer Interaction program at Utrecht University. This study is carried out by me, Ingi Páll Eiríksson (i.p.eiriksson@students.uu.nl) as part of my master thesis under the supervision of Dr. A.A. Akdag (a.a.akdag@uu.nl).

In this study, we are investigating variations in respiration patterns in response to commonly known user interface features. You will be asked to create an imaginary persona for yourself and navigate through an online shopping environment making decisions based on that persona. After the experiment, you will be asked to fill out a short questionnaire. You will be wearing a breathing belt throughout the experiment, which is estimated to take 10-15 minutes.

Data and privacy

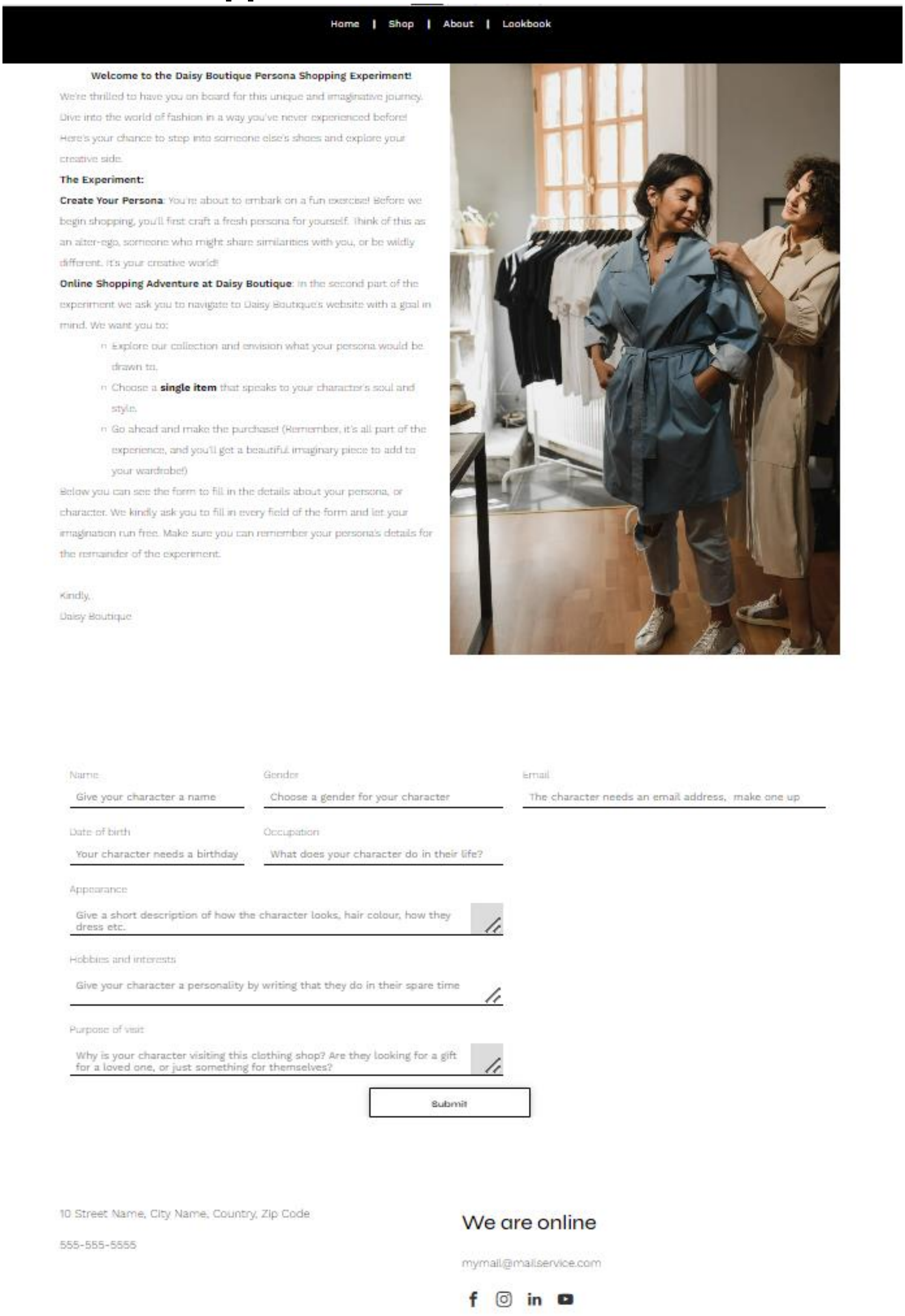
In the course of this study, we will not collect any data that directly identifies you. All information you offer will be kept anonymous and treated with strict confidentiality. Your decision to engage in this study is entirely voluntary. By choosing to participate, you consent to the gathering and utilization of your data for the purposes of this research. However, you are under no obligation to provide any reasoning if you decline participation. At any given point, should you wish to cease participation, you are entirely within your rights to do so, without any obligation to inform us of your reasons. Even after having participated, you retain the right to retract your consent and discontinue further involvement. If you decide to withdraw after providing data any personally identifiable data will be erased.

This study has been allowed to proceed by the Research Institute of Information and Computing Sciences following an Ethics and Privacy Quick Scan. Should you harbor any concerns about the manner in which this research is conducted, we kindly request you to reach out to ics-ethics@uu.nl. For queries or concerns related to personal data processing, the Faculty of Sciences Privacy Officer stands ready to assist at privacy-beta@uu.nl. This officer is also your point of contact to exercise any rights under the GDPR. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information at www.uu.nl/en/organisation/privacy.

Please read the statements below and click check the box below to confirm you have read and understood the statements and upon doing so agree to participate in the project.

· I confirm that the research project has been explained to me. I have had the opportunity to ask questions about the project and have had these answered satisfactorily.
· I confirm that I am 18 years of age or over.
· I consent to the material I contribute being used to generate insights for the research project.
· I understand that my participation in this research is voluntary and that I may withdraw from the study at any time without providing a reason, and that if I withdraw any personal data already collected from me will be erased.
· I understand that the information/data acquired will be securely stored by researchers, but that appropriately anonymized data may in the future be made available to others for research purposes only.
· I understand that I can request any of the data collected from/by me to be deleted.

Appendix 2 – The website



Home | Shop | About | Lookbook

Welcome to the Daisy Boutique Persona Shopping Experiment!

We're thrilled to have you on board for this unique and imaginative journey. Dive into the world of fashion in a way you've never experienced before! Here's your chance to step into someone else's shoes and explore your creative side.

The Experiment:

Create Your Persona: You're about to embark on a fun exercise! Before we begin shopping, you'll first craft a fresh persona for yourself. Think of this as an alter-ego, someone who might share similarities with you, or be wildly different. It's your creative world!

Online Shopping Adventure at Daisy Boutique: In the second part of the experiment we ask you to navigate to Daisy Boutique's website with a goal in mind. We want you to:

- Explore our collection and envision what your persona would be drawn to.
- Choose a **single item** that speaks to your character's soul and style.
- Go ahead and make the purchase! (Remember, it's all part of the experience, and you'll get a beautiful imaginary piece to add to your wardrobe!)

Below you can see the form to fill in the details about your persona, or character. We kindly ask you to fill in every field of the form and let your imagination run free. Make sure you can remember your persona's details for the remainder of the experiment.

Kindly,
Daisy Boutique

Name
Give your character a name

Gender
Choose a gender for your character

Email
The character needs an email address, make one up

Date of birth
Your character needs a birthday

Occupation
What does your character do in their life?

Appearance
Give a short description of how the character looks, hair colour, how they dress etc.

Hobbies and interests
Give your character a personality by writing that they do in their spare time

Purpose of visit
Why is your character visiting this clothing shop? Are they looking for a gift for a loved one, or just something for themselves?

Submit

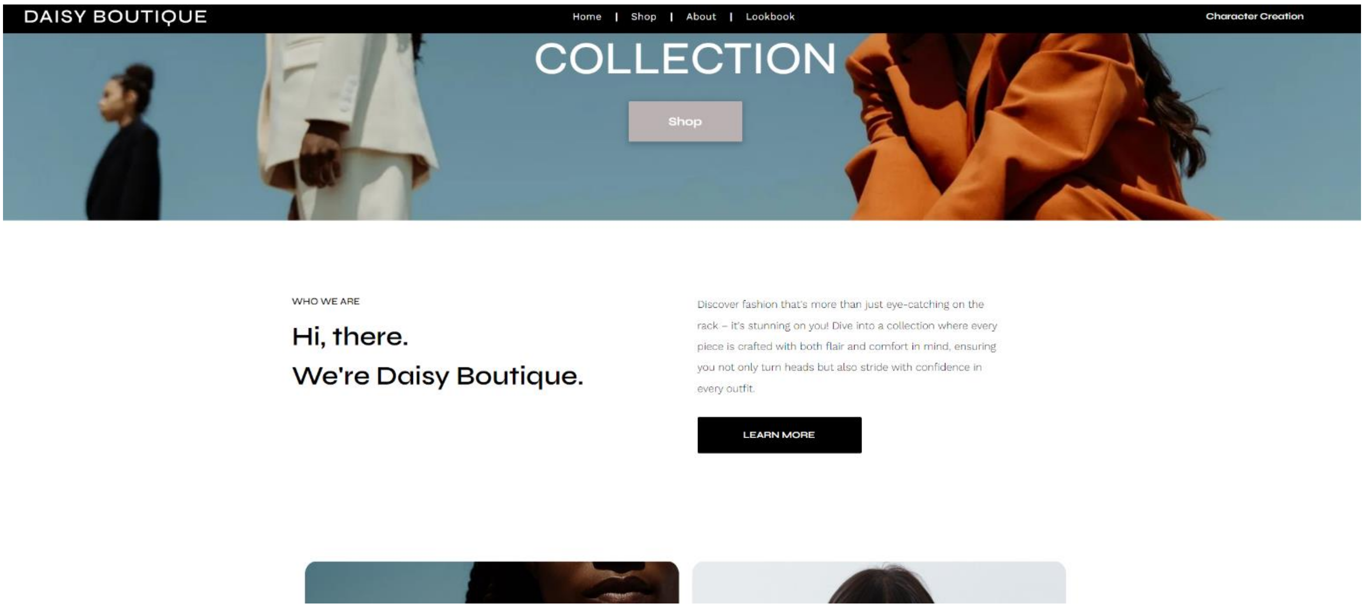
10 Street Name, City Name, Country, Zip Code

555-555-5555

We are online

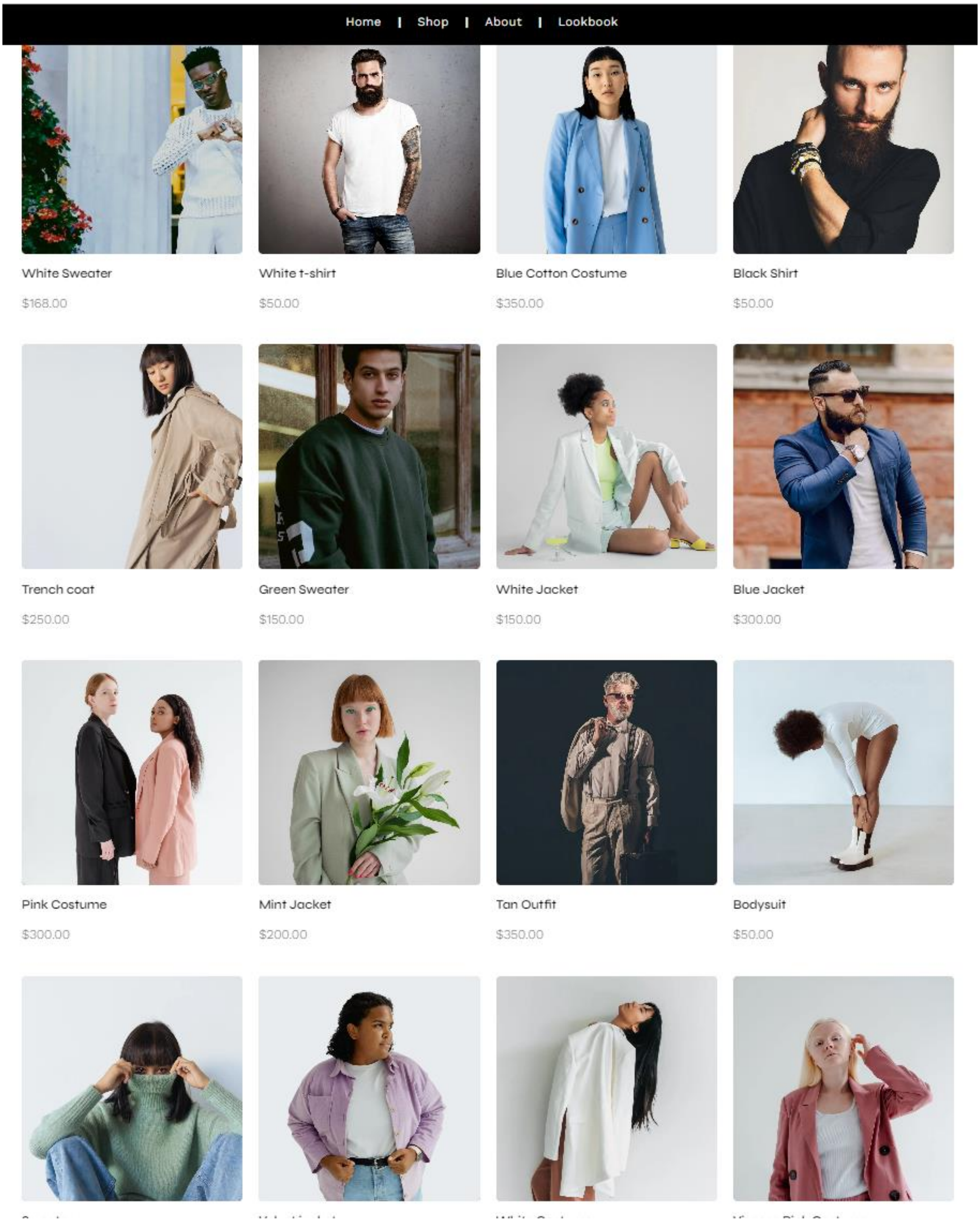
mymail@mailservice.com

for character creation.

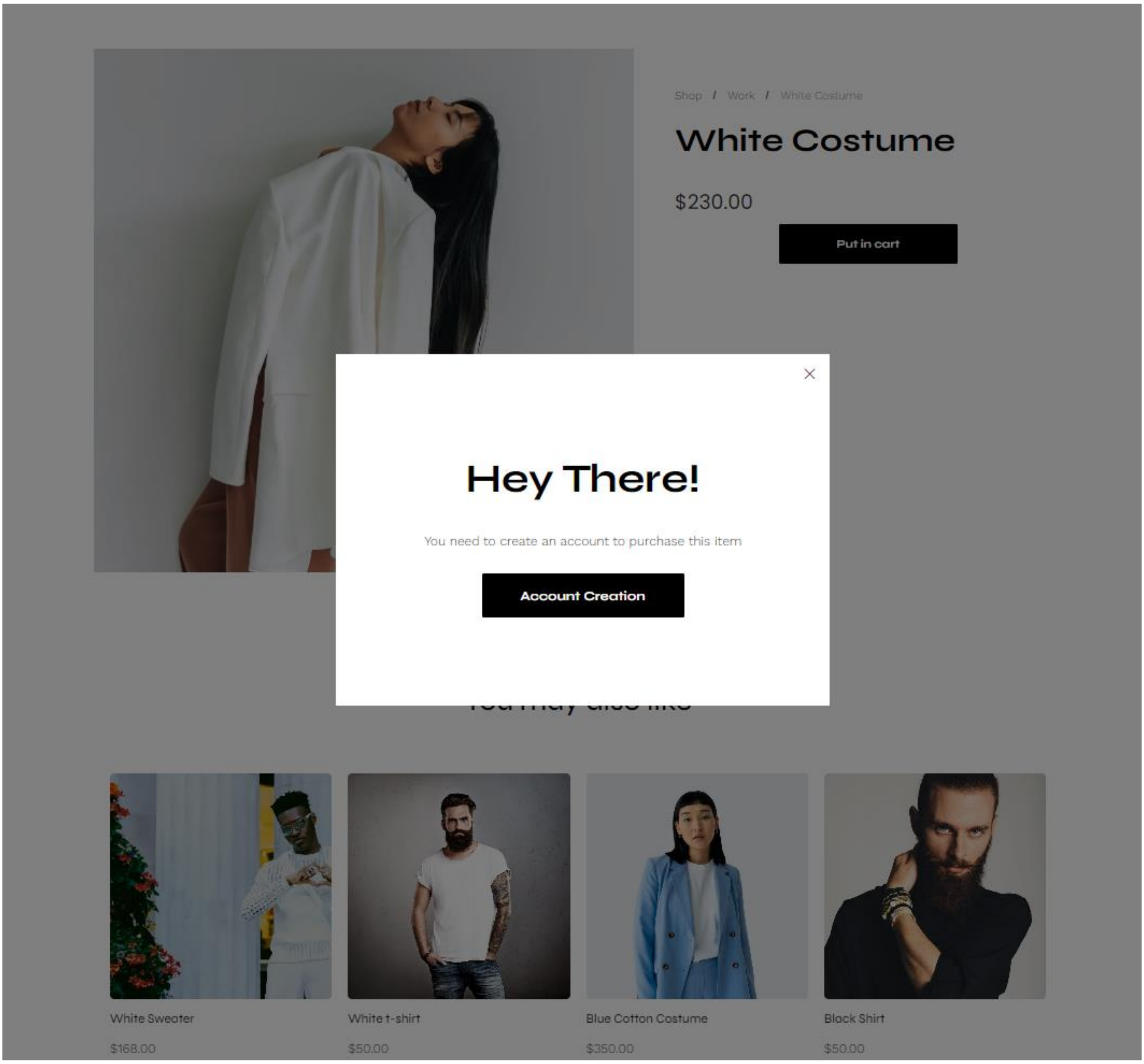


The opening page to the website, every image is clickable and brings the user to the shop.

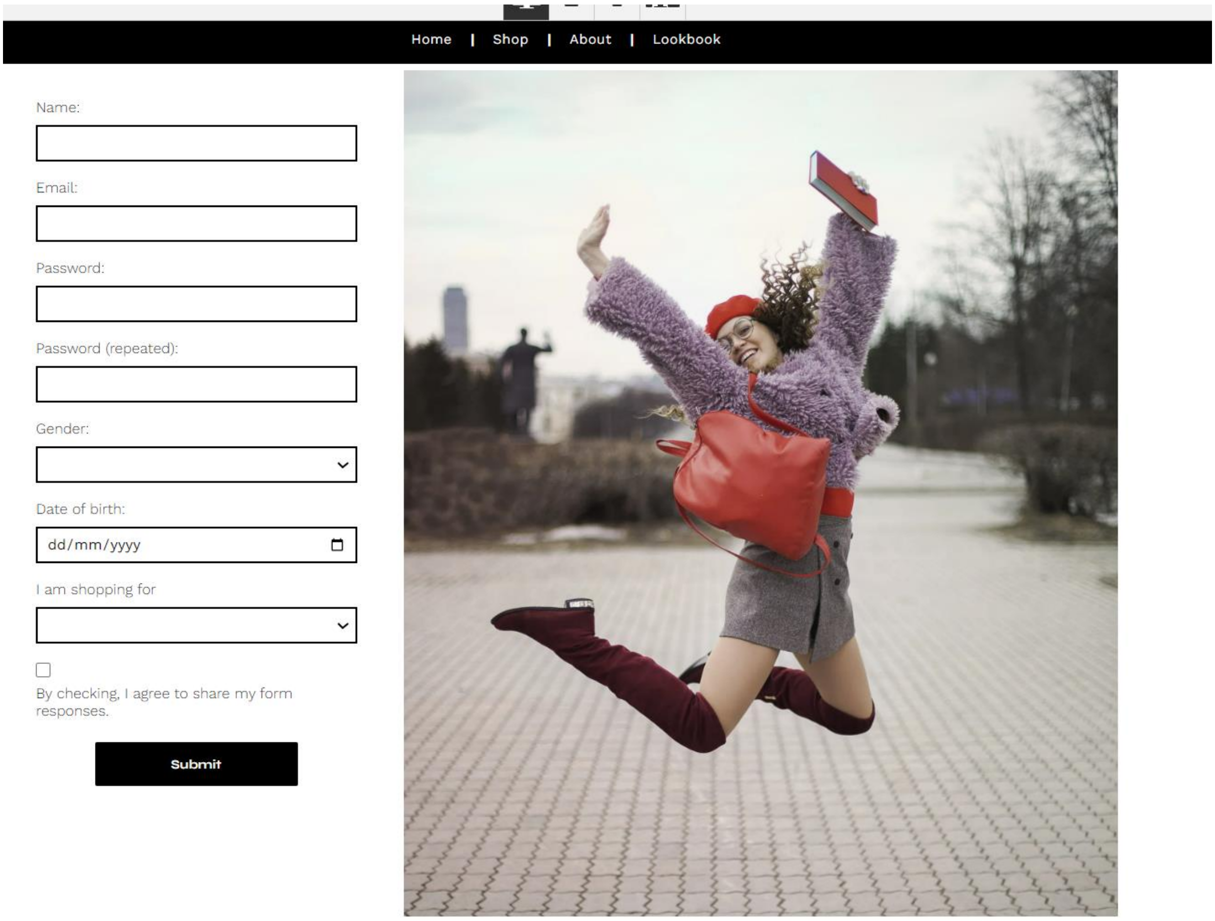
The lower half of the opening page with changed proportions.



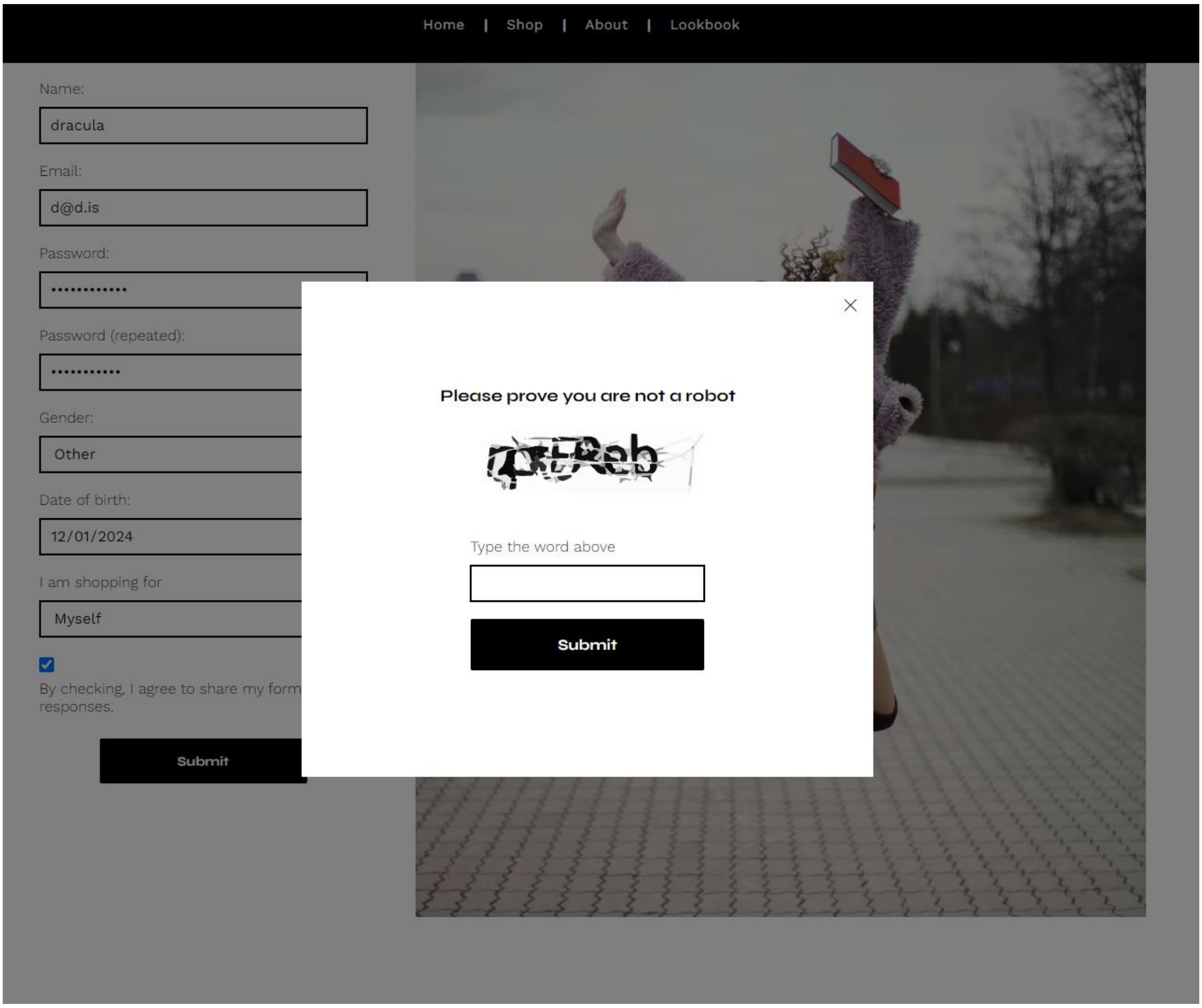
A selection of items available for purchase.



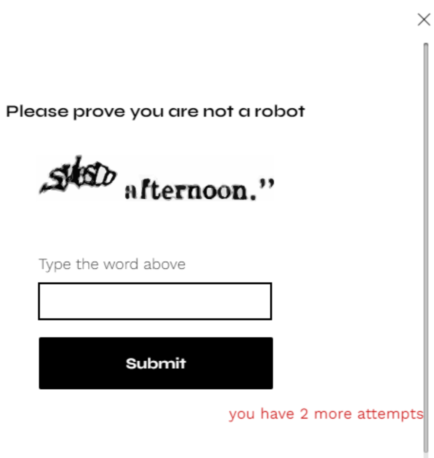
A popup window sending the user to the account creation site as they try to purchase an item.



The account creation page before anything is entered, clicking submit starts the account creation task by prompting a popup window

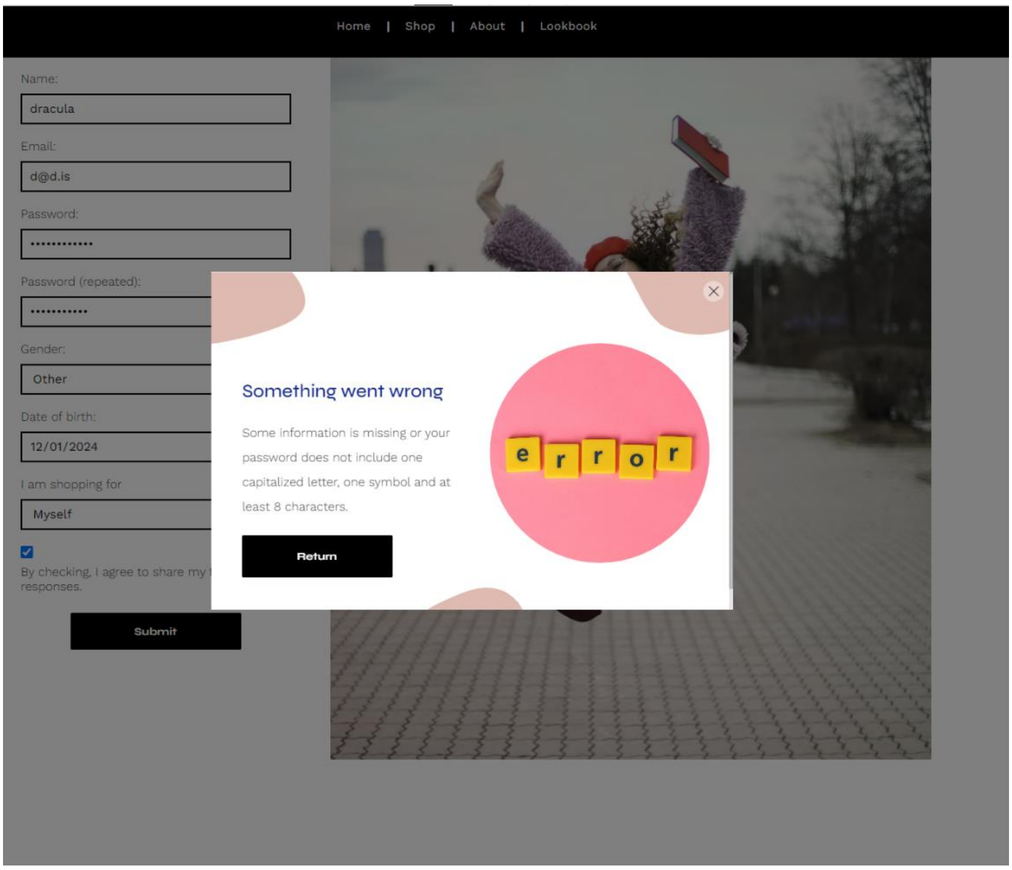


The first popup CAPTCHA window, this one is wrong no matter what is written.

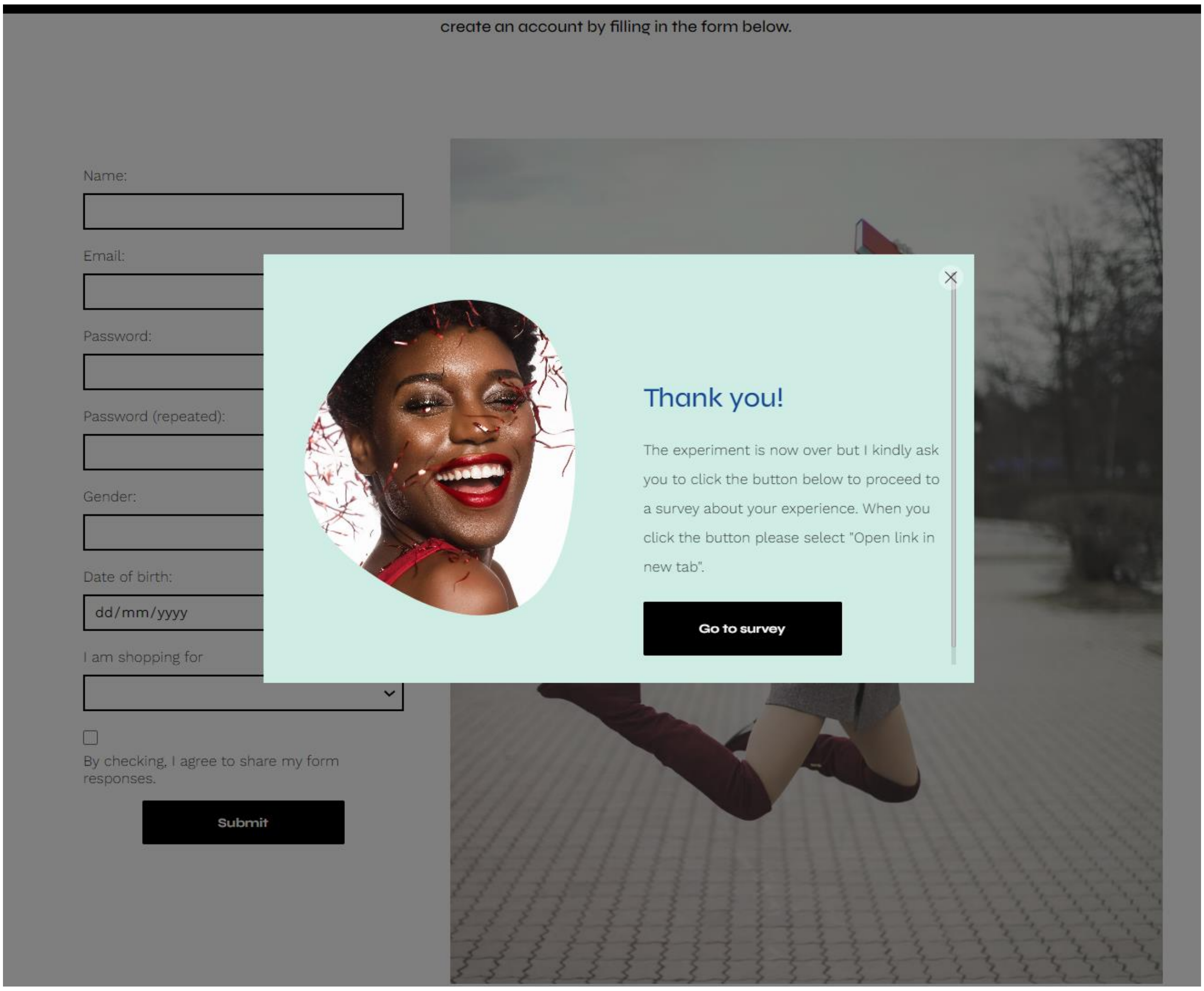


The second CAPTCHA window that appears after the first one, this one has red letters notifying of remaining attempts, it is also wrong no matter what is written.

The third and last CAPCHA popup, this one prompts an error message and wipes out everything the user has written in the form.



The error message that pops up after the CAPCHA is done.



The popup that appears after all the information has been entered again, pressing the button brings the user to the survey and begins the fourth task.

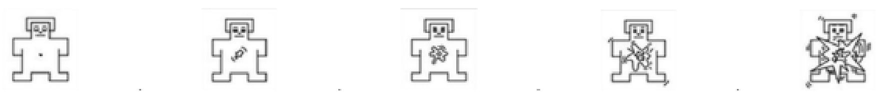
Appendix 3 – The survey

Thank you for your participation!

The online shopping part of the experiment is now over. I kindly ask you to answer a few questions about your experience with it and that you keep your breathing belt on as it has been throughout the study.

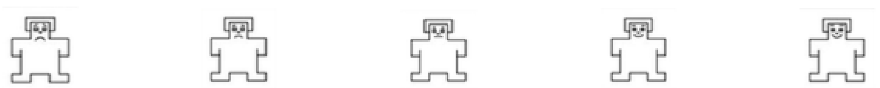
Arousal (low to high)
The first picture shows a person that is **very calm or even sleepy**. Relevant states could include tranquillity, relaxation, boredom, idleness or laziness. The last picture shows a person that is **bursting with arousal**. Relevant states could include excitation, rage, anger, euphoria, agitation or excitement.

Please indicate how you felt after the **"account creation"** (not to be confused with the persona or character creation in the beginning of the experiment). Note that you can mark in between two figures.



Valence (low to high)
The first picture shows a person that is **distressed**. Relevant experiences could include panic, despair, irritation or defeat. The last picture shows a person that is **clearly pleased**. Relevant experiences could include fun, happiness, satisfaction, delight or repose.

Please indicate how you felt after the **"account creation"** process that you just completed (not to be confused with the persona or character creation in the beginning of the experiment). Note that you can mark in between two figures.



→

The next few questions are about the Character or persona creation page in the beginning of the experiment.

Please choose the option that best describes your state **during the persona creation process** in the beginning of the experiment.

Completely calm	Mostly calm	Slightly calm	Neutral	Slightly anxious	Mostly anxious	Extremely anxious
○	○	○	○	○	○	○

Please choose the option that best describes your state **during the persona creation proces**s in the beginning of the experiment.

Completely pleased	Mostly pleased	Slightly pleased	Neutral	Slightly annoyed	Mostly annoyed	Extremely annoyed
○	○	○	○	○	○	○

← →

Please choose the option that best describes your **current state**.

Completely pleased	Mostly pleased	Slightly pleased	Neutral	Slightly annoyed	Mostly annoyed	Extremely annoyed
○	○	○	○	○	○	○

Please choose the option that best describes your **current state**.

Completely calm	Mostly calm	Slightly calm	Neutral	Slightly anxious	Mostly anxious	Extremely anxious
○	○	○	○	○	○	○

Is there anything you would like to add, any suggestions or complaints?

← →

Appendix 4 – Tables with descriptive statistics and normalcy tests

index	mean	std	min	25%	50%	75%	max
Q1 T3, A, normalized	3.75	1.48	1.00	2.50	3.63	4.75	6.25
Q1 T3, V, normalized	4.17	1.44	1.75	3.25	4.00	4.75	7.00
Q3, T1,2, A	2.78	1.56	1.00	2.00	2.00	4.75	5.00
Q3, T1,2, V	2.67	1.41	1.00	2.00	2.00	3.00	5.00
Q5, T4, V	2.83	1.42	1.00	2.00	2.50	4.00	6.00
Q5, T4, A	2.44	1.29	1.00	2.00	2.00	3.00	5.00

Table A.1. Dessriptive statistics for the six quantitative questions, Question 1 is related to the account creation process, task 3, and measures arousal, question 2 is also to task 3, and measures valence. Question 3 is related to both the character creation and the shopping experience, tasks 1 and 2, and measures arousal, question 4 also relates to tasks 1 and 2 and measures valence. Question 5 is related to the end of the experiment (survey), task 4, and measured valence, and question 6 relates to task 4 and measures arousal.

Question	Statistic (W)	P-Value
Arousal, Q1(T3)	0.957	0.544
Arousal, Q3(T1,2)	0.763	0.000
Arousal, Q6(T4)	0.847	0.008
Valence, Q2(T3)	0.960	0.603
Valence, Q4 (T1,2)	0.809	0.002
Valence , Q5(T4)	0.916	0.110

Table A.2. Results from the Shapiro-Wilk test for the six questions for tasks T1 (character creation), T2 (shopping), T3 (account creation), and T4 (survey). Note that T1 and T2 are combined in T1,2

Task	mean	median	std
RR (Character creation)	15.621	15.666	1.350
RR (Shopping)	15.984	16.010	1.627
RR (Account creation)	15.577	15.664	1.500
RR (Survey)	15.641	15.804	2.245
E (Everything)	15.421	15.353	1.545

*Table A.3. Descriptive statistics for average respiration rate (RR)
For the four tasks and the experiment as a whole.*

Task	W	p-value	Kurtosis	Skewness
RR (Character creation)	0.955	0.513	-1.11	-0.204
RR (Shopping)	0.987	0.995	-0.2	0.127
RR (Account creation)	0.953	0.478	-0.177	-0.551
RR (Survey)	0.967	0.735	-0.466	-0.432
RR (Everything)	0.937	0.254	-0.08	-0.691

Table A.4. Results of the tests for normalcy for average inhales and exhales per minute (RR).

	Distance between troughs			**Distance between peaks**		
Task	**Mean**	**Median**	**Std**	**Mean**	**Median**	**Std**
Task 1 (Character creation)	2963.02	2871.88	280.94	3066.57	2973.39	357.06
Task 2 (Shopping)	2833.21	2876.42	202.70	2996.33	2885.21	514.22
Task 3 (Account creation)	2988.22	2890.90	321.93	2995.96	2898.91	338.29
Task 4 (Survey)	2977.33	2904.12	403.82	3041.95	2890.08	546.54
Everything	3060.62	2946.61	367.29	3239.47	3128.87	480.81

Table A.5. Descriptive statistics for the average distance between peaks and troughs.

Task	W	p-value	Skewness	Kurtosis
DBT (Character creation)	0.930	0.196	0.906	0.430
DBT (Shopping)	0.845	0.007	-1.402	1.375
DBT (Account creation)	0.911	0.088	0.910	0.238
DBT (Survey)	0.886	0.033	1.372	2.448
Everything	0.771	0.001	2.013	3.962
DBP (Character creation)	0.897	0.050	0.825	-0.204
DBP (Shopping)	0.905	0.071	1.063	0.483
DBP (Account creation)	0.909	0.084	0.899	-0.053
DBP (Survey)	0.930	0.193	0.655	-0.431
DBP (Everything)	0.895	0.047	0.654	-0.805

Table A.6. Results of the tests for normalcy for the average distance between peaks (DBP) and troughs (DBT)

Task	**mean**	**median**	**std**
PH (Character creation)	0.48	0.44	0.21
PH (Shopping)	0.42	0.42	0.21
PH (Account creation)	0.43	0.48	0.21
PH (Survey)	0.40	0.42	0.20
PH (Everything)	0.44	0.43	0.21

Table A.7. Descriptive statistics for the average peak height (PH) for each task and the experiment as a whole

Tasks	**W**	**p-value**	**Skewness**	**Kurtosis**
PH (Character creation)	0.951	0.441	0.285	-1.103
PH (Shopping)	0.979	0.935	-0.046	-0.907
PH (Account creation)	0.966	0.727	-0.052	-0.718
PH (Survey)	0.983	0.975	-0.109	-0.788
PH (Everything)	0.973	0.857	0.069	-0.915

Table A.8. Statistical tests for normalcy for PH

	Average sharpness of peaks			**Average sharpness of troughs**		
Task	**mean**	**median**	**std**	**mean**	**median**	**std**
Task 1 (Character creation)	0.22	0.20	0.10	-0.19	-0.17	0.10
Task 2 (Shopping)	0.18	0.18	0.08	-0.15	-0.14	0.11
Task 3 (Account creation)	0.21	0.18	0.10	-0.16	-0.15	0.10
Task 4 (Survey)	0.15	0.16	0.07	-0.13	-0.12	0.11
Everything	0.19	0.19	0.08	-0.16	-0.15	0.10

Table A.9. Descriptive statistics for average sharpness of peaks and troughs multiplied by 1000.

Task	**W**	**p-value**	**Skewness**	**Kurtosis**
PS (Character creation)	0.883	0.029	1.380	2.495
PS (Shopping)	0.932	0.208	0.650	-0.145
PS (Account creation)	0.883	0.029	1.030	0.227
PS (Survey)	0.953	0.470	0.215	-1.061
PS (Everything)	0.927	0.169	0.996	0.815
TS (Character creation)	0.932	0.208	-0.853	0.709
TS (Shopping)	0.934	0.225	-0.891	1.024
TS (Account creation)	0.886	0.032	-1.024	0.911
TS (Survey)	0.844	0.007	-1.648	3.630
PS (Everything)	0.925	0.158	-1.020	1.276

Table A.10.Tests for normalcy for average peak sharpness (PS) and trout sharpness (TS).

Appendix 5 – F and p matrices for the survey questions.

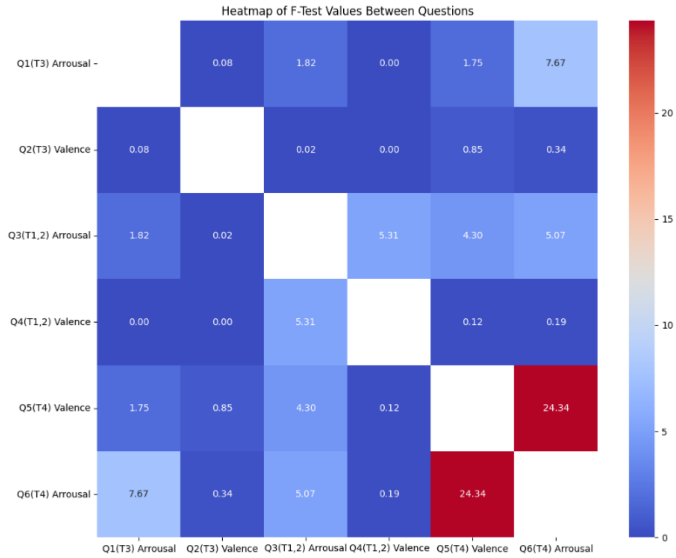


Table A.11. A heat map of the F values from the F test of R2 values from the linear regression of the questions

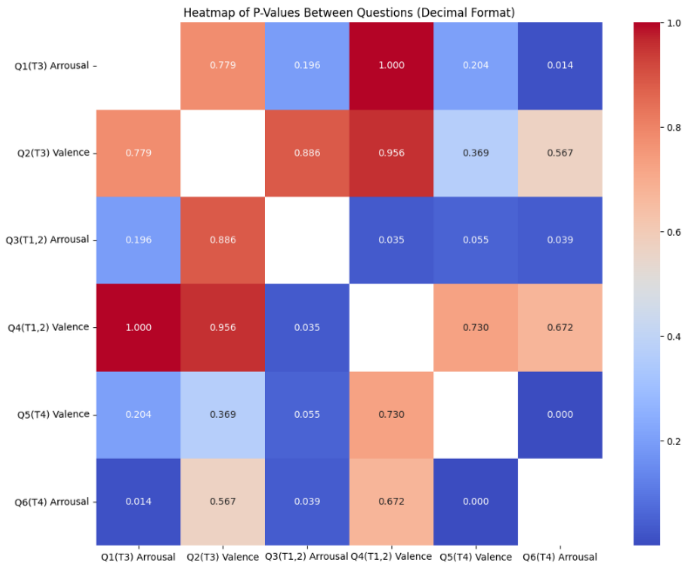


Table A.12. A heatmap with the P values from the F test of R2 values from the linear regression of the questions.

Appendix 6 – IPM and EPM and difference

IPM1	IPM2	IPM3	IPM4	IPM5	EPM1	EPM2	EPM3	EPM4	EPM5
16.40465	13.73997	15.69975	17.12818	14.43426	15.12303	16.1777	16.90742	16.31255	15.50652
16.5438	17.34912	16.40273	16.29482	16.51592	18.10454	17.34912	16.98854	15.94058	16.51592
15.95701	15.86491	15.13576	14.45052	15.73874	16.53576	15.50434	15.59442	15.89557	16.31391
15.56875	16.21183	15.29843	14.12191	15.04895	15.56875	15.8669	14.87347	13.75981	15.04895
12.77465	16.80229	17.41992	19.10017	13.9758	14.03117	16.47917	15.96826	18.39276	16.41453
16.48526	19.0408	18.47047	18.01061	17.59704	16.80228	19.85105	17.15115	17.34355	17.82858
16.94635	18.07574	16.44316	19.13392	17.39965	16.65417	15.66564	16.99127	16.74218	16.54192
17.66667	15.16894	16.57244	12.0947	16.35669	17.82583	15.54817	15.01877	12.54265	15.89809
16.91548	18.02975	17.30474	17.6764	17.34996	16.91548	15.40041	18.29359	15.1512	16.56132
16.52808	17.57741	16.10198	17.85714	16.8702	15.95152	17.24576	14.49178	17.85714	16.8702
14.25482	15.04453	15.53281	16.0307	14.45622	15.87879	15.04453	15.53281	15.05914	15.34858
14.4426	15.04054	15.37925	14.56647	14.24143	16.52875	16.92061	13.67044	15.74119	15.51179
14.87021	17.9095	17.43054	18.98247	16.81289	14.34845	18.70547	16.26851	18.98247	16.99271
16.25711	16.09406	14.97144	16.199	14.86499	17.8203	16.28566	16.63494	15.92901	16.1576
13.1078	14.02091	13.67594	10.38445	12.39245	14.66825	11.95901	11.00746	11.26823	11.4041
12.58596	11.22457	12.84613	14.63396	11.48963	14.22761	15.90147	12.84613	13.8835	13.93118
14.05099	14.10267	15.45993	13.8089	13.98554	15.29078	16.27231	14.32872	16.378	15.40653
13.95198	12.08231	13.28522	11.47753	12.61758	14.80272	15.85803	14.76136	13.937	14.77181

Table A.13. A table showing the values for inhales and exhales per minute for every task

1	2	3	4	5
1.281613	-2.43774	-1.20767	0.815627	-1.07226
-1.56074	0	-0.58581	0.354235	0
-0.57875	0.360566	-0.45866	-1.44505	-0.57517
0	0.344933	0.424956	0.3621	0
-1.25652	0.323121	1.45166	0.707414	-2.43873
-0.31702	-0.81025	1.319319	0.667059	-0.23154
0.292178	2.410098	-0.54811	2.391741	0.857729
-0.15916	-0.37922	1.553666	-0.44795	0.458599
0	2.629338	-0.98884	2.525199	0.788634
0.576561	0.331649	1.610198	0	0
-1.62397	0	0	0.971558	-0.89236
-2.08615	-1.88007	1.708805	-1.17472	-1.27036
0.521762	-0.79598	1.162036	0	-0.17982
-1.56318	-0.1916	-1.66349	0.269983	-1.29261
-1.56045	2.061898	2.668475	-0.88378	0.988355
-1.64165	-4.6769	0	0.75046	-2.44155
-1.23979	-2.16964	1.131215	-2.5691	-1.42099
-0.85073	-3.77572	-1.47614	-2.45947	-2.15422

Table A.14. A table showing the differences between imp and emp for each task, the numbers in red are not fully explainable by the study setup, a difference of more than 2.

Appendix 7 – Correlations with p values for questions and breathing data

	Q1A	IEPM3	mDBT3	mDBP3	stdDBT3	stdDBP3	mPH3	mPS3	mTS3
Q1A	1.00, -	-0.09, 0.714	0.07, 0.786	-0.01, 0.976	-0.04, 0.881	0.02, 0.928	-0.27, 0.28	0.36, 0.142	0.13, 0.593

	Q2V	IEPM3	mDBT3	mDBP3	stdDBT3	stdDBP3	mPH3	mPS3	mTS3
Q2V	1.00, -	0.14, 0.589	-0.25, 0.322	-0.04, 0.866	-0.26, 0.303	-0.11, 0.65	-0.09, 0.717	-0.43, 0.0748	0.08, 0.765

CC	Q3A	IEPM1	mDBT1	mDBP1	stdDBT1	stdDBP1	mPH1	mPS1	mTS1
Q3A	1.00, -	-0.35, 0.157	0.47, 0.0477	0.28, 0.261	0.47, 0.0483	0.36, 0.141	-0.24, 0.335	0.40, 0.0989	-0.15, 0.547

S	Q3A	IEPM2	mDBT2	mDBP2	stdDBT2	stdDBP2	mPH2	mPS2	mTS2
Q3A	1.00, -	0.05, 0.843	-0.04, 0.871	0.02, 0.925	0.13, 0.595	0.08, 0.746	-0.20, 0.433	0.46, 0.0568	-0.15, 0.56

CC	Q4V	IEPM1	mDBT1	mDBP1	stdDBT1	stdDBP1	mPH1	mPS1	mTS1
Q4V	1.00, -	-0.27, 0.283	0.47, 0.0474	-0.08, 0.765	0.42, 0.0799	-0.02, 0.948	0.07, 0.79	0.02, 0.932	-0.22, 0.373

S	Q4V	IEPM2	mDBT2	mDBP2	stdDBT2	stdDBP2	mPH2	mPS2	mTS2
Q4V	1.00, -	0.14, 0.582	-0.15, 0.551	0.26, 0.296	-0.21, 0.408	0.34, 0.173	0.10, 0.685	0.19, 0.448	-0.43, 0.078

	Q5V	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
Q5V	1.00, -	0.58, 0.0122	-0.55, 0.0184	-0.51, 0.0316	-0.47, 0.0498	-0.50, 0.0348	0.14, 0.575	0.29, 0.24	-0.14, 0.591

	Q6A	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
Q6A	1.00, -	0.40, 0.104	-0.36, 0.137	-0.38, 0.121	-0.26, 0.29	-0.35, 0.155	-0.10, 0.693	0.24, 0.33	0.10, 0.7

Table A.15. A correlation table for each question and it's correlation with each breathing measurement and the p values on the form [correlation, p-value].

Appendix 8 – Median of breathing values aggregated by survey responses and std of means.

				1 median					
Answer for Q1A	Number of Participants	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
1	1	14.89433	3313.287	2765.489	2221.853	1670.244	0.259134	0.000133	-0.00013
2	1	17.79917	2536.554	2777.552	1521.939	1708.276	0.806728	0.000227	-0.00039
3	4	16.05339	2738.344	3099.493	1666.663	2123.695	0.526311	0.00018	-0.00019
4	3	15.53281	2958.664	2908.691	1837.564	1833.186	0.550117	0.000128	-0.0001
5	2	14.91086	3134.936	3011.15	1920.928	1883.503	0.244527	0.000278	-0.00016
6	3	12.84613	3357.49	3494.784	2172.692	2503.097	0.323674	0.00018	-6.5E-05
7	3	16.71721	2859.011	2633.428	1737.271	1633.676	0.472051	0.000419	-0.00022
8	1	15.36509	2888.444	3052.639	1724.436	1904.313	0.329553	0.000162	-0.00017
Kruskal-Wallis p-value	NaN	0.465321	0.383379	0.33666	0.540547	0.231433	0.330374	0.524931	0.222357
				2 median					
Answer for Q2V	Number of Participants	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
2	1	16.71721	2801.268	2879.663	1737.271	1783.335	0.77262	0.000419	-0.0004
3	3	12.84613	3357.49	3494.784	2221.853	2503.097	0.259134	0.000138	-0.00013
4	3	16.84953	2859.011	2608.514	1716.993	1635.986	0.472051	0.000301	-9.6E-05
5	3	15.80319	2746.229	3063.836	1681.437	1844.147	0.575951	0.000176	-0.00018
6	4	15.54624	2887.024	2814.115	1730.607	1716.714	0.244527	0.000243	-0.00016
7	1	16.30358	2683.361	2976.073	1651.889	1911.176	0.338689	0.000197	-0.0002
8	2	15.35869	3090.049	3124.402	1978.234	2039.289	0.532832	0.000104	-0.00012
9	1	15.36509	2888.444	3052.639	1724.436	1904.313	0.329553	0.000162	-0.00017
Kruskal-Wallis p-value	NaN	0.131549	0.173663	0.312546	0.082284	0.370592	0.370066	0.039801	0.779336
				3 median					
Answer for Q3A	Number of Participants	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
1	3	17.0387	2677.26	2729.051	1667.207	1668.205	0.520096	0.000194	-0.00013
2	9	15.56875	2866.046	2998.201	1740.442	1854.583	0.566613	0.000173	-0.00016
4	1	16.80026	2823.172	2775.79	1713.081	1732.201	0.843886	0.000502	-0.00038
5	5	14.60933	3071.095	3061.168	1961.543	2047.857	0.367342	0.00023	-0.00018
Kruskal-Wallis p-value	NaN	0.203841	0.293166	0.377789	0.262909	0.242231	0.36612	0.148562	0.647921

Table A.16. Table created to see if there was a big difference between using the median instead of the mean when aggregating the responses to the first 3 questions. There was not a big difference.

				4 median					
Answer for Q4V	Number of Participants	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
1	3	16.24638	2748.912	2948.587	1671.02	1789.917	0.342588	0.000192	-0.00013
2	8	16.06472	2851.105	3011.894	1750.651	1969.867	0.46334	0.000202	-0.00014
3	3	15.56875	2839.743	2998.201	1698.595	1854.583	0.650205	0.000219	-0.00022
5	4	15.18659	3083.802	2968.907	2046.858	1960.12	0.387952	0.000227	-0.0002
Kruskal-Wallis p-value	NaN	0.6539	0.252911	0.885779	0.214518	0.929043	0.43082	0.98858	0.487568
				5 median					
Answer for Q5V	Number of Participants	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
1	3	15.54492	3061.38	2911.617	1885.811	1787.022	0.459129	0.000099	-0.00011
2	6	13.90036	3100.537	3510.32	1931.12	2624.306	0.327412	0.000163	-8.1E-05
3	3	15.15383	2998.581	3255.412	1869.165	1968.838	0.274957	0.000181	-0.00006
4	4	16.98742	2738.483	2730.182	1645.332	1674.033	0.366879	0.000203	-0.00019
5	1	17.93805	2442.316	2352.842	1447.33	1403.594	0.679783	0.000231	-0.00014
6	1	18.74646	2682.192	2459.72	1628.437	1477.653	0.502043	0.00007	-0.00015
Kruskal-Wallis p-value	NaN	0.230697	0.054611	0.267886	0.067465	0.271562	0.787911	0.303543	0.135162
				6 median					
Answer for Q6A	Number of Participants	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
1	4	15.16137	3118.796	3127.437	1895.07	2119.268	0.527439	0.000104	-0.00012
2	8	15.34938	2946.564	3083.514	1795.619	1862.388	0.327412	0.000147	-0.0001
3	2	14.25171	3419.31	3231.221	2308.216	2244.531	0.172856	0.000177	-1.5E-05
4	2	18.46026	2450.203	2375.113	1463.982	1443.213	0.544246	0.000258	-0.00017
5	2	16.95975	2760.974	2838.4	1675.24	1730.209	0.413547	0.00012	-0.00015
Kruskal-Wallis p-value	NaN	0.216488	0.09822	0.182448	0.119699	0.316354	0.179143	0.253556	0.256811

Table A.17. Table created to see if there was a big difference between using the median instead of the mean when aggregating the responses to the last 3 questions. There was not a big difference.

				Std 1					
Answer for Q6A	Number of Participants	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
3	4	1.180033	284.6219	296.6733	260.4805	339.3484	0.109531	0.000042	0.000033
4	3	0.830097	120.236	183.7233	30.19487	131.408	0.061153	0.000107	0.000007
5	2	0.545911	322.5094	226.9412	290.8413	223.9416	0.037925	0.000027	0.000012
6	3	3.022517	567.8289	615.478	377.476	622.3649	0.231282	0.000067	0.00006
7	3	0.57411	35.92176	149.8741	14.8574	101.1295	0.35776	0.000192	0.000194
				std 2					
3	3	1.351879	268.1266	523.9712	96.07572	603.3616	0.23393	2.63E-05	0.000066
4	3	1.143613	145.8446	175.3155	103.4615	128.9385	0.115026	1.01E-04	0.00008
5	3	0.80643	224.202	166.9562	126.7182	313.6528	0.068496	3.02E-05	0.000052
6	4	1.398108	339.7157	227.5349	253.008	180.8316	0.324499	8.24E-05	0.000158
8	2	1.888539	278.1751	603.3043	260.9498	605.3261	0.046316	7.28E-07	0.000013
				std 3					
1	3	1.456478	306.5124	309.4671	294.9125	321.2152	0.177846	0.000016	0.000072
2	9	1.338067	173.5595	281.5272	213.5713	296.5092	0.231208	0.000077	0.000125
5	5	1.206243	369.0431	500.3612	650.2376	770.985	0.095276	0.000049	0.000037
				std 4					
1	3	1.638924	101.5725	464.2667	41.22558	550.2411	0.133704	0.000027	0.000051
2	8	1.275647	229.337	329.5859	241.1843	331.4212	0.241918	0.000132	0.000114
3	3	1.772134	490.1254	601.7915	869.7802	1031.758	0.172335	0.000073	0.000167
5	4	1.266415	211.3602	206.0321	391.1749	213.532	0.205625	0.000077	0.000036
				std 5					
1	3	0.929402	154.4113	287.323	95.40878	392.9248	0.03982	0.000011	0.000021
2	6	2.407967	497.7023	639.4733	499.7564	629.6172	0.28023	0.00007	0.000173
3	3	1.906017	284.2357	466.7514	172.6227	442.9504	0.233247	0.000071	0.000046
4	4	1.710445	180.4164	337.9348	112.7808	209.774	0.16692	0.000067	0.00003
				std 6					
1	4	1.718498	245.7487	687.0833	229.5498	565.4401	0.147726	0.000083	0.000184
2	8	1.693247	193.3017	357.3591	145.9909	470.9662	0.205398	0.000061	0.000073
3	2	4.844205	1071.876	1072.984	1006.893	1107.596	0.144393	0.000005	0.000004
4	2	0.738516	11.15444	31.49631	23.5496	56.03031	0.191679	0.000039	0.000043
5	2	2.526788	111.4152	535.5342	66.18965	357.1677	0.125153	0.000071	0.000005

Table A.18. Table of standard deviations of the means of the breathing data aggregated by question responses.

Appendix 9 – F and P values for the correlation and regression analysis

Character creation and shopping (table A.19)

R 1 2	IEPM1	mDBT1	mDBP1	stdDBT1	stdDBP1	mPH1	mPS1	mTS1
IEPM2	0.230, p: 0.0442, F: 4.77	0.000, p: 0.993, F: 7.33e-05	0.218, p: 0.0506, F: 4.47	0.005, p: 0.777, F: 0.0829	0.115, p: 0.169, F: 2.08	0.001, p: 0.907, F: 0.014	0.317, p: 0.015, F: 7.43	0.008, p: 0.73, F: 0.123
mDBT2	0.025, p: 0.53, F: 0.412	0.005, p: 0.79, F: 0.0731	0.053, p: 0.358, F: 0.897	0.009, p: 0.705, F: 0.149	0.046, p: 0.393, F: 0.771	0.036, p: 0.45, F: 0.601	0.109, p: 0.182, F: 1.95	0.074, p: 0.274, F: 1.28
mDBP2	0.326, p: 0.0133, F: 7.75	0.052, p: 0.362, F: 0.882	0.144, p: 0.12, F: 2.7	0.018, p: 0.592, F: 0.299	0.074, p: 0.275, F: 1.28	0.000, p: 0.961, F: 0.00242	0.338, p: 0.0114, F: 8.16	0.019, p: 0.585, F: 0.311
stdDBT2	0.014, p: 0.641, F: 0.226	0.002, p: 0.845, F: 0.0395	0.051, p: 0.368, F: 0.858	0.002, p: 0.878, F: 0.0242	0.071, p: 0.286, F: 1.22	0.049, p: 0.375, F: 0.833	0.000, p: 0.93, F: 0.00786	0.099, p: 0.204, F: 1.75
stdDBP2	0.260, p: 0.0305, F: 5.63	0.055, p: 0.348, F: 0.934	0.100, p: 0.2, F: 1.78	0.021, p: 0.562, F: 0.35	0.059, p: 0.333, F: 0.997	0.000, p: 0.942, F: 0.00551	0.261, p: 0.0303, F: 5.65	0.010, p: 0.7, F: 0.154
mPH2	0.000, p: 0.972, F: 0.00126	0.004, p: 0.804, F: 0.0635	0.005, p: 0.775, F: 0.0842	0.004, p: 0.812, F: 0.0584	0.002, p: 0.853, F: 0.0357	0.949, p: 8.79e-12, F: 299	0.008, p: 0.731, F: 0.122	0.469, p: 0.00172, F: 14.1
mPS2	0.003, p: 0.819, F: 0.0542	0.011, p: 0.68, F: 0.176	0.019, p: 0.581, F: 0.318	0.004, p: 0.793, F: 0.071	0.026, p: 0.522, F: 0.429	0.001, p: 0.926, F: 0.00883	0.728, p: 6.72e-06, F: 42.9	0.120, p: 0.16, F: 2.17
mTS2	0.006, p: 0.763, F: 0.0945	0.019, p: 0.586, F: 0.308	0.043, p: 0.408, F: 0.723	0.024, p: 0.541, F: 0.39	0.021, p: 0.568, F: 0.34	0.330, p: 0.0126, F: 7.89	0.067, p: 0.301, F: 1.14	0.796, p: 6.51e-07, F: 62.4

Character creation and account creation (table A.20)

R 1 3	IEPM1	mDBT1	mDBP1	stdDBT1	stdDBP1	mPH1	mPS1	mTS1
IEPM3	0.326, p: 0.0133, F: 7.75	0.000, p: 0.97, F: 0.00142	0.233, p: 0.0424, F: 4.86	0.007, p: 0.744, F: 0.111	0.087, p: 0.236, F: 1.52	0.027, p: 0.513, F: 0.447	0.200, p: 0.0625, F: 4.01	0.084, p: 0.242, F: 1.48
mDBT3	0.445, p: 0.0025, F: 12.8	0.009, p: 0.702, F: 0.151	0.417, p: 0.00379, F: 11.5	0.001, p: 0.896, F: 0.0177	0.207, p: 0.0579, F: 4.17	0.053, p: 0.357, F: 0.898	0.090, p: 0.226, F: 1.58	0.110, p: 0.18, F: 1.97
mDBP3	0.205, p: 0.0593, F: 4.12	0.010, p: 0.692, F: 0.162	0.104, p: 0.191, F: 1.86	0.029, p: 0.497, F: 0.482	0.020, p: 0.574, F: 0.329	0.055, p: 0.35, F: 0.925	0.178, p: 0.0808, F: 3.47	0.000, p: 0.969, F: 0.00153
stdDBT3	0.488, p: 0.00125, F: 15.3	0.031, p: 0.482, F: 0.519	0.445, p: 0.0025, F: 12.8	0.009, p: 0.713, F: 0.141	0.229, p: 0.0443, F: 4.76	0.032, p: 0.475, F: 0.534	0.093, p: 0.219, F: 1.64	0.096, p: 0.21, F: 1.71
stdDBP3	0.187, p: 0.0734, F: 3.67	0.011, p: 0.682, F: 0.174	0.083, p: 0.248, F: 1.44	0.023, p: 0.548, F: 0.377	0.016, p: 0.618, F: 0.258	0.035, p: 0.458, F: 0.578	0.144, p: 0.12, F: 2.7	0.001, p: 0.885, F: 0.0214
mPH3	0.001, p: 0.882, F: 0.0227	0.009, p: 0.703, F: 0.151	0.010, p: 0.693, F: 0.161	0.011, p: 0.673, F: 0.185	0.005, p: 0.79, F: 0.0731	0.887, p: 5.7e-09, F: 125	0.024, p: 0.536, F: 0.4	0.467, p: 0.00176, F: 14
mPS3	0.024, p: 0.542, F: 0.387	0.013, p: 0.659, F: 0.203	0.081, p: 0.253, F: 1.41	0.012, p: 0.672, F: 0.186	0.078, p: 0.26, F: 1.36	0.017, p: 0.609, F: 0.273	0.563, p: 0.000337, F: 20.6	0.114, p: 0.171, F: 2.05
mTS3	0.092, p: 0.222, F: 1.62	0.003, p: 0.826, F: 0.0499	0.144, p: 0.12, F: 2.7	0.005, p: 0.786, F: 0.0764	0.112, p: 0.174, F: 2.03	0.434, p: 0.00294, F: 12.3	0.266, p: 0.0284, F: 5.8	0.793, p: 7.33e-07, F: 61.3

Character creation and survey (table A.21)

R 1 4	IEPM1	mDBT1	mDBP1	stdDBT1	stdDBP1	mPH1	mPS1	mTS1
IEPM4	0.018, p: 0.597, F: 0.291	0.158, p: 0.102, F: 3	0.040, p: 0.424, F: 0.674	0.200, p: 0.063, F: 3.99	0.002, p: 0.863, F: 0.0306	0.025, p: 0.528, F: 0.416	0.360, p: 0.00851, F: 8.99	0.159, p: 0.102, F: 3.01
mDBT4	0.113, p: 0.172, F: 2.04	0.064, p: 0.311, F: 1.09	0.160, p: 0.1, F: 3.05	0.080, p: 0.256, F: 1.39	0.066, p: 0.302, F: 1.14	0.012, p: 0.665, F: 0.195	0.301, p: 0.0184, F: 6.89	0.082, p: 0.249, F: 1.43
mDBP4	0.088, p: 0.231, F: 1.55	0.029, p: 0.5, F: 0.477	0.113, p: 0.173, F: 2.03	0.068, p: 0.297, F: 1.16	0.024, p: 0.54, F: 0.393	0.048, p: 0.382, F: 0.808	0.419, p: 0.00369, F: 11.5	0.190, p: 0.0703, F: 3.76
stdDBT4	0.120, p: 0.159, F: 2.19	0.051, p: 0.367, F: 0.861	0.187, p: 0.0728, F: 3.69	0.059, p: 0.33, F: 1.01	0.089, p: 0.229, F: 1.57	0.033, p: 0.472, F: 0.544	0.191, p: 0.0696, F: 3.78	0.088, p: 0.233, F: 1.54
stdDBP4	0.069, p: 0.291, F: 1.19	0.034, p: 0.462, F: 0.567	0.092, p: 0.221, F: 1.62	0.059, p: 0.33, F: 1.01	0.025, p: 0.529, F: 0.413	0.115, p: 0.168, F: 2.08	0.303, p: 0.0179, F: 6.95	0.214, p: 0.0532, F: 4.36
mPH4	0.000, p: 0.962, F: 0.00229	0.014, p: 0.635, F: 0.234	0.007, p: 0.747, F: 0.108	0.015, p: 0.629, F: 0.243	0.002, p: 0.863, F: 0.0309	0.882, p: 7.55e-09, F: 120	0.005, p: 0.775, F: 0.0842	0.450, p: 0.0023, F: 13.1
mPS4	0.036, p: 0.454, F: 0.59	0.002, p: 0.859, F: 0.0325	0.098, p: 0.205, F: 1.74	0.005, p: 0.782, F: 0.079	0.110, p: 0.178, F: 1.98	0.008, p: 0.731, F: 0.123	0.481, p: 0.00141, F: 14.8	0.205, p: 0.0594, F: 4.12
mTS4	0.034, p: 0.467, F: 0.556	0.002, p: 0.867, F: 0.0288	0.074, p: 0.276, F: 1.27	0.004, p: 0.794, F: 0.0705	0.035, p: 0.46, F: 0.574	0.296, p: 0.0196, F: 6.72	0.015, p: 0.63, F: 0.242	0.681, p: 2.47e-05, F: 34.2

Shopping and account creation (table A.22)

R 2 3	IEPM3	mDBT3	mDBP3	stdDBT3	stdDBP3	mPH3	mPS3	mTS3
IEPM2	0.687, p: 2.11e-05, F: 35.2	0.485, p: 0.00133, F: 15.1	0.567, p: 0.000311, F: 20.9	0.518, p: 0.000759, F: 17.2	0.521, p: 0.00072, F: 17.4	0.036, p: 0.451, F: 0.597	0.291, p: 0.0209, F: 6.56	0.096, p: 0.21, F: 1.71
mDBT2	0.215, p: 0.0525, F: 4.39	0.095, p: 0.214, F: 1.68	0.151, p: 0.111, F: 2.84	0.104, p: 0.191, F: 1.86	0.113, p: 0.173, F: 2.04	0.000, p: 0.938, F: 0.00626	0.185, p: 0.0749, F: 3.63	0.003, p: 0.831, F: 0.0469
mDBP2	0.529, p: 0.000629, F: 17.9	0.331, p: 0.0126, F: 7.9	0.587, p: 0.000208, F: 22.8	0.406, p: 0.00447, F: 10.9	0.582, p: 0.00023, F: 22.3	0.033, p: 0.473, F: 0.54	0.241, p: 0.0385, F: 5.08	0.087, p: 0.235, F: 1.52
stdDBT2	0.108, p: 0.183, F: 1.94	0.061, p: 0.323, F: 1.04	0.067, p: 0.301, F: 1.14	0.075, p: 0.272, F: 1.29	0.060, p: 0.328, F: 1.02	0.003, p: 0.826, F: 0.05	0.088, p: 0.232, F: 1.54	0.019, p: 0.59, F: 0.303
stdDBP2	0.391, p: 0.00555, F: 10.3	0.215, p: 0.0527, F: 4.38	0.472, p: 0.00163, F: 14.3	0.290, p: 0.0212, F: 6.53	0.491, p: 0.0012, F: 15.4	0.033, p: 0.473, F: 0.539	0.186, p: 0.0736, F: 3.67	0.062, p: 0.321, F: 1.05
mPH2	0.078, p: 0.262, F: 1.35	0.104, p: 0.192, F: 1.85	0.020, p: 0.575, F: 0.328	0.070, p: 0.29, F: 1.2	0.008, p: 0.719, F: 0.134	0.960, p: 1.44e-12, F: 379	0.020, p: 0.574, F: 0.33	0.438, p: 0.00276, F: 12.5
mPS2	0.089, p: 0.23, F: 1.55	0.018, p: 0.592, F: 0.299	0.126, p: 0.149, F: 2.3	0.025, p: 0.534, F: 0.403	0.120, p: 0.158, F: 2.19	0.007, p: 0.738, F: 0.115	0.794, p: 7.04e-07, F: 61.7	0.221, p: 0.0487, F: 4.55
mTS2	0.156, p: 0.105, F: 2.95	0.187, p: 0.0729, F: 3.69	0.019, p: 0.585, F: 0.31	0.172, p: 0.087, F: 3.33	0.029, p: 0.499, F: 0.479	0.390, p: 0.00562, F: 10.2	0.147, p: 0.117, F: 2.75	0.728, p: 6.67e-06, F: 42.9

Shopping and survey (Table A.23)

R 2 4	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
IEPM2	0.612, p: 0.000126, F: 25.2	0.618, p: 0.000109, F: 25.9	0.620, p: 0.000104, F: 26.2	0.557, p: 0.000374, F: 20.1	0.586, p: 0.000215, F: 22.6	0.023, p: 0.55, F: 0.373	0.217, p: 0.0511, F: 4.45	0.061, p: 0.323, F: 1.04
mDBT2	0.261, p: 0.0304, F: 5.64	0.198, p: 0.0641, F: 3.96	0.209, p: 0.0564, F: 4.23	0.157, p: 0.103, F: 2.98	0.221, p: 0.0491, F: 4.53	0.003, p: 0.83, F: 0.0479	0.096, p: 0.211, F: 1.7	0.012, p: 0.67, F: 0.189
mDBP2	0.288, p: 0.0216, F: 6.48	0.294, p: 0.0202, F: 6.65	0.427, p: 0.00326, F: 11.9	0.244, p: 0.0373, F: 5.16	0.419, p: 0.00366, F: 11.6	0.016, p: 0.622, F: 0.253	0.154, p: 0.107, F: 2.92	0.045, p: 0.396, F: 0.76
stdDBT2	0.152, p: 0.11, F: 2.87	0.127, p: 0.147, F: 2.32	0.098, p: 0.205, F: 1.75	0.128, p: 0.144, F: 2.36	0.144, p: 0.12, F: 2.69	0.006, p: 0.766, F: 0.0913	0.054, p: 0.354, F: 0.909	0.001, p: 0.884, F: 0.0221
stdDBP2	0.203, p: 0.0603, F: 4.09	0.209, p: 0.0566, F: 4.22	0.299, p: 0.0188, F: 6.83	0.168, p: 0.091, F: 3.23	0.327, p: 0.0132, F: 7.77	0.017, p: 0.605, F: 0.278	0.125, p: 0.15, F: 2.28	0.034, p: 0.463, F: 0.564
mPH2	0.072, p: 0.283, F: 1.24	0.038, p: 0.435, F: 0.64	0.090, p: 0.228, F: 1.57	0.073, p: 0.28, F: 1.25	0.198, p: 0.064, F: 3.96	0.972, p: 7.35e-14, F: 557	0.008, p: 0.729, F: 0.124	0.327, p: 0.0132, F: 7.77
mPS2	0.275, p: 0.0253, F: 6.08	0.215, p: 0.0527, F: 4.38	0.286, p: 0.0221, F: 6.42	0.136, p: 0.132, F: 2.52	0.218, p: 0.0507, F: 4.46	0.000, p: 0.97, F: 0.00143	0.709, p: 1.16e-05, F: 39.1	0.022, p: 0.553, F: 0.368
mTS2	0.235, p: 0.0414, F: 4.92	0.150, p: 0.112, F: 2.82	0.219, p: 0.0501, F: 4.49	0.161, p: 0.0983, F: 3.08	0.241, p: 0.0385, F: 5.08	0.395, p: 0.00525, F: 10.4	0.293, p: 0.0203, F: 6.64	0.904, p: 1.52e-09, F: 150

Account creation and survey (Table A.24)

R 3 4	IEPM4	mDBT4	mDBP4	stdDBT4	stdDBP4	mPH4	mPS4	mTS4
IEPM3	0.549, p: 0.000439, F: 19.4	0.557, p: 0.000378, F: 20.1	0.642, p: 6.43e-05, F: 28.7	0.536, p: 0.000556, F: 18.4	0.616, p: 0.000115, F: 25.6	0.131, p: 0.14, F: 2.41	0.097, p: 0.209, F: 1.72	0.219, p: 0.0503, F: 4.48
mDBT3	0.458, p: 0.00205, F: 13.5	0.515, p: 0.000805, F: 17	0.568, p: 0.000305, F: 21	0.538, p: 0.000533, F: 18.6	0.567, p: 0.000312, F: 20.9	0.146, p: 0.118, F: 2.73	0.039, p: 0.432, F: 0.65	0.275, p: 0.0255, F: 6.07
mDBP3	0.361, p: 0.00832, F: 9.05	0.439, p: 0.00274, F: 12.5	0.422, p: 0.0035, F: 11.7	0.386, p: 0.00595, F: 10	0.314, p: 0.0155, F: 7.33	0.002, p: 0.876, F: 0.025	0.110, p: 0.18, F: 1.97	0.045, p: 0.399, F: 0.75
stdDBT3	0.408, p: 0.00431, F: 11	0.431, p: 0.00309, F: 12.1	0.569, p: 0.000296, F: 21.2	0.444, p: 0.00252, F: 12.8	0.568, p: 0.000302, F: 21.1	0.106, p: 0.188, F: 1.89	0.056, p: 0.343, F: 0.956	0.269, p: 0.0274, F: 5.89
stdDBP3	0.330, p: 0.0126, F: 7.89	0.422, p: 0.00352, F: 11.7	0.370, p: 0.00737, F: 9.41	0.385, p: 0.00597, F: 10	0.308, p: 0.0167, F: 7.13	0.000, p: 0.943, F: 0.00534	0.102, p: 0.197, F: 1.82	0.060, p: 0.329, F: 1.01
mPH3	0.132, p: 0.138, F: 2.44	0.087, p: 0.234, F: 1.53	0.145, p: 0.119, F: 2.71	0.126, p: 0.148, F: 2.31	0.268, p: 0.0277, F: 5.86	0.967, p: 2.98e-13, F: 465	0.000, p: 0.995, F: 4.02e-05	0.358, p: 0.00874, F: 8.91
mPS3	0.393, p: 0.00538, F: 10.4	0.336, p: 0.0117, F: 8.1	0.401, p: 0.00477, F: 10.7	0.259, p: 0.0312, F: 5.58	0.335, p: 0.0118, F: 8.07	0.022, p: 0.559, F: 0.356	0.543, p: 0.00049, F: 19	0.054, p: 0.353, F: 0.916
mTS3	0.278, p: 0.0246, F: 6.15	0.250, p: 0.0348, F: 5.32	0.278, p: 0.0246, F: 6.15	0.249, p: 0.0349, F: 5.31	0.316, p: 0.0153, F: 7.38	0.463, p: 0.00187, F: 13.8	0.320, p: 0.0144, F: 7.54	0.603, p: 0.000149, F: 24.4

Appendix 10 – Differences between valence and arousal correlations with breathing data

	Q1, T3, A	RR3	mDBT3	mDBP3	stdDBT3	stdDBP3	PH3	PS3	TS3	average	combined
Q1	1.0***	-0.09	0.07	-0.01	-0.04	0.02	-0.27	0.36	0.13		
Q2V	1.0***	0.14	-0.25	-0.04	-0.26	-0.11	-0.09	-0.43	0.08		
		0.23	0.32	0.03	0.22	0.13	0.18	0.79	0.05	0.24375	1.95
Q3A	1.0***	-0.35	0.47	0.28	0.47	0.36	-0.24	0.4	-0.15		
Q4V	1.0***	-0.27	0.47	-0.08	0.42	-0.02	0.07	0.02	-0.22		
		0.08	0	0.36	0.05	0.38	0.31	0.38	0.07	0.20375	1.63
Q5V	1.0***	0.58	-0.55	-0.51	-0.47	-0.5	0.14	0.29	-0.14		
Q6A	1.0***	0.4	-0.36	-0.38	-0.26	-0.35	-0.1	0.24	0.1		
		0.18	0.19	0.13	0.21	0.15	0.24	0.05	0.24	0.17375	1.39

s

Q3A	1.0***	0.05	-0.04	0.02	0.13	0.08	-0.2	0.46	-0.15		
Q4V	1.0***	0.14	-0.15	0.26	-0.21	0.34	0.1	0.19	-0.43		
		0.09	0.11	0.24	0.34	0.26	0.3	0.27	0.28	0.23625	1.89

Table A.25. Differences between the correlations of each question pair with the breathing data.

Appendix 11 - Correlations of survey responses to the study as a whole and regression analysis results of the same

Below are 6 tables, one for each question marked in the format Q1V meaning question one Valence. They include the correlations with various breathing measurements , p values, R^2 and F statistics from the regression analysis

Account creation Arrousal (Table A.26)

Question	Predictor	Coefficient	Corr P-value	R-squared	F-statistic	Correlation
Q1A	IEPM5	0.166237426	0.606097364	0.016998829	0.276684574	0.130379556
Q1A	mDBT5	0.000262813	0.846928457	0.002399984	0.038492128	0.048989634
Q1A	mDBP5	-0.000556931	0.590786314	0.018469472	0.301072202	-0.135902437
Q1A	stdDBT5	0.000522408	0.595845005	0.017975108	0.292866029	0.13407128
Q1A	stdDBP5	-0.000268342	0.685508852	0.01051853	0.170085535	-0.102559887
Q1A	mPH5	-3.063097043	0.195095765	0.102566611	1.828621265	-0.320260224
Q1A	mPS5	10063.92499	0.088042764	0.171001919	3.300406549	0.413523782
Q1A	mTS5	3379.387621	0.495335327	0.029531775	0.486887046	0.171848117

Account creation valence (Table A.27)

Question	Predictor	Coefficient	Corr P-value	R-squared	F-statistic	Correlation
Q2V	IEPM5	0.118722514	0.711383847	0.008788533	0.141863297	0.093747176
Q2V	mDBT5	-0.001297065	0.330370849	0.059255485	1.007805774	-0.243424495
Q2V	mDBP5	0.0004379	0.670904728	0.011574198	0.18735566	0.107583447
Q2V	stdDBT5	-0.001360102	0.152702053	0.123504428	2.254513222	-0.351431968
Q2V	stdDBP5	0.000351163	0.592926454	0.018259294	0.297582351	0.135126956
Q2V	mPH5	-1.009137042	0.674835958	0.011284274	0.182608992	-0.106227463
Q2V	mPS5	-10990.91938	0.057990465	0.206739316	4.169914271	-0.454685953
Q2V	mTS5	-890.1697856	0.857491552	0.002077055	0.033302056	-0.045574723

Character creation arousal (Table A.28)

Question	Predictor	Coefficient	Corr P-value	R-squared	F-statistic	Correlation
Q3A	IEPM5	-0.028855546	0.910073404	0.000822252	0.013166851	-0.028674929
Q3A	mDBT5	0.000693006	0.516358904	0.026790128	0.44044153	0.163676901
Q3A	mDBP5	0.000493227	0.545778474	0.023255774	0.380951709	0.152498439
Q3A	stdDBT5	0.000631939	0.41334577	0.04222655	0.705411918	0.205490998
Q3A	stdDBP5	0.000572577	0.265298019	0.076882966	1.332580172	0.277277778
Q3A	mPH5	-1.571213018	0.407200941	0.043325078	0.724594354	-0.20814677
Q3A	mPS5	8340.387608	0.071776377	0.188548634	3.717755945	0.434221872
Q3A	mTS5	-1875.115782	0.632978345	0.014596692	0.237006581	-0.12081677

Character creation valence (Table A.29)

Question	Predictor	Coefficient	Corr P-value	R-squared	F-statistic	Correlation
Q4V	IEPM5	0.075566	0.744624	0.006818	0.109844	0.082574
Q4V	mDBT5	-0.0003	0.762082	0.005893	0.094842	-0.07676
Q4V	mDBP5	0.000189	0.799709	0.004142	0.066555	0.064362
Q4V	stdDBT5	-5.00E-05	0.943815	0.00032	0.005125	-0.01789
Q4V	stdDBP5	0.000197	0.679126	0.010973	0.177512	0.104751
Q4V	mPH5	0.658366	0.705019	0.009198	0.148531	0.095905
Q4V	mPS5	3744.145	0.393047	0.045945	0.770518	0.214347
Q4V	mTS5	-3924.22	0.263948	0.077301	1.340437	-0.27803

Survey Valence (Table A.30)

Question	Predictor	Coefficient	Corr P-value	R-squared	F-statistic	Correlation
Q5V	IEPM5	0.43145	0.050137	0.219052	4.487932	0.46803
Q5V	mDBT5	-0.00179	0.054008	0.21277	4.324423	-0.46127
Q5V	mDBP5	-0.00066	0.373712	0.049738	0.837456	-0.22302
Q5V	stdDBT5	-0.00125	0.064687	0.197438	3.936153	-0.44434
Q5V	stdDBP5	-6.52E-05	0.892067	0.001187	0.019007	-0.03445
Q5V	mPH5	0.560409	0.749216	0.006568	0.105779	0.081042
Q5V	mPS5	7727.529	0.068239	0.192873	3.823408	0.439174
Q5V	mTS5	-3919.91	0.268129	0.076014	1.316272	-0.27571

Survey Arousal (Table A.31)

Question	Predictor	Coefficient	Corr P-value	R-squared	F-statistic	Correlation
Q6A	IEPM5	0.216244	0.300631	0.066742	1.144237	0.258344
Q6A	mDBT5	-0.00072	0.412928	0.0423	0.706702	-0.20567
Q6A	mDBP5	-0.00017	0.804469	0.003944	0.063358	-0.0628
Q6A	stdDBT5	-0.00045	0.483325	0.031184	0.51501	-0.17659
Q6A	stdDBP5	0.000157	0.717889	0.008381	0.13523	0.091548
Q6A	mPH5	-0.97264	0.53938	0.023996	0.393372	-0.15491
Q6A	mPS5	6821.478	0.077209	0.182293	3.566907	0.426958
Q6A	mTS5	-673.013	0.837231	0.002718	0.043602	-0.05213

Appendix 12 – Statistical tests for the OLS assumptions

Below are tables of the results of the various statistical rest conducted to test validity of the OLS model starting with the questions, then the question-metrics interaction and finally the list of the metrics for every task. The values that do not reach a significance level of 0.05 are marked with red and values that are concidered on the line are marked orange.

	Variable Pair	Shapiro-Wilk Statistic	Shapiro-Wilk P-value	Durbin-Watson	Rainbow Statistic	Rainbow P-value	Breusch-Pagan Statistic	Breusch-Pagan P-value
0	Q1A vs Q2V	0.969	0.784	1.966	3.193	0.070	2.008	0.156
1	Q1A vs Q3A	0.846	0.007	2.565	1.452	0.319	0.104	0.747
2	Q1A vs Q4V	0.809	0.002	2.324	0.950	0.540	1.379	0.240
3	Q1A vs Q5V	0.964	0.680	2.019	1.893	0.206	0.016	0.898
4	Q1A vs Q6A	0.954	0.489	2.363	4.119	0.038	0.147	0.701
5	Q2V vs Q3A	0.778	0.001	2.735	1.183	0.422	0.030	0.863
6	Q2V vs Q4V	0.817	0.003	2.321	0.912	0.562	0.030	0.862
7	Q2V vs Q5V	0.929	0.188	2.036	1.199	0.415	0.092	0.762
8	Q2V vs Q6A	0.892	0.042	2.128	4.145	0.037	3.092	0.079
9	Q3A vs Q4V	0.966	0.725	2.407	3.842	0.045	2.328	0.127
10	Q3A vs Q5V	0.943	0.320	1.429	3.378	0.061	1.932	0.165
11	Q3A vs Q6A	0.920	0.128	1.542	2.158	0.161	0.209	0.647
12	Q4V vs Q5V	0.949	0.406	1.863	1.596	0.275	0.083	0.773
13	Q4V vs Q6A	0.886	0.033	2.203	1.652	0.260	0.131	0.717
14	Q5V vs Q6A	0.973	0.851	2.816	1.420	0.329	0.998	0.318

Table A.32. The results between questions.

Question	Predictor	Shapiro-Wilk P-value	Breusch-Pagan P-value	Durbin-Watson	Rainbow P-value
Q3A	mDBT1	0.186	0.991	2.212	0.255
Q3A	stdDBT1	0.039	0.549	2.389	0.351
Q5V	IEPM4	0.650	0.277	2.197	0.477
Q5V	mDBT4	0.954	0.998	2.388	0.080
Q5V	mDBP4	0.564	0.140	2.341	0.525
Q5V	stdDBT4	0.866	0.777	2.453	0.099
Q5V	stdDBP4	0.477	0.083	2.548	0.692
Q5V	mPH4	0.544	0.165	2.033	0.559
Q5V	mPS4	0.089	0.212	1.636	0.081
Q5V	mTS4	0.309	0.435	1.876	0.558

Table A.33. The results from the relationships of between the questions and the respiration metrics, only the ones that are used in the paper, the rest can be requested.

	Variable Pair	S-W Statistic	S-W P-value	Durbin-Watson	Ljung-Box Statistic	Ljung-Box P-value	Breusch-Pagan Statistic	Breusch-Pagan P-value	Rainbow Statistic	Rainbow P-value
0	IEPM1 vs IEPM2	0.918	0.119	1.510	20.301	0.259	0.610	0.435	0.948	0.541
1	IEPM1 vs IEPM3	0.960	0.596	1.840	15.51	0.559	4.835	0.028	1.287	0.378
2	IEPM1 vs IEPM4	0.964	0.686	1.901	9.085	0.938	1.794	0.180	1.791	0.227
3	IEPM2 vs IEPM3	0.965	0.695	2.105	10.005	0.903	0.056	0.813	0.715	0.687
4	IEPM2 vs IEPM4	0.980	0.951	2.923	23.023	0.148	0.058	0.810	1.089	0.466
5	IEPM3 vs IEPM4	0.971	0.809	2.259	6.592	0.988	0.019	0.890	0.376	0.913
6	mDBT1 vs mDBT2	0.841	0.006	2.126	12.214	0.787	0.049	0.825	0.529	0.816
7	mDBT1 vs mDBT3	0.897	0.051	1.321	18.01	0.388	0.725	0.395	2.560	0.114
8	mDBT1 vs mDBT4	0.894	0.045	1.631	14.482	0.633	0.514	0.473	3.656	0.051
9	mDBT2 vs mDBT3	0.955	0.508	1.353	16.535	0.486	2.315	0.128	2.167	0.160
10	mDBT2 vs mDBT4	0.901	0.059	2.256	9.231	0.933	1.026	0.311	3.389	0.061
11	mDBT3 vs mDBT4	0.947	0.383	2.104	12.921	0.741	1.457	0.227	1.021	0.501
12	mDBP1 vs mDBP2	0.897	0.052	1.471	28.134	0.043	0.834	0.361	4.682	0.027
13	mDBP1 vs mDBP3	0.974	0.867	1.851	12.204	0.788	4.185	0.041	2.783	0.096
14	mDBP1 vs mDBP4	0.983	0.978	1.583	9.345	0.929	5.294	0.021	1.722	0.243
15	mDBP2 vs mDBP3	0.905	0.070	2.138	23.911	0.122	1.045	0.307	0.966	0.531
16	mDBP2 vs mDBP4	0.971	0.818	2.088	21.811	0.192	8.606	0.003	4.066	0.039
17	mDBP3 vs mDBP4	0.927	0.172	2.308	11.728	0.816	0.000	0.995	0.913	0.561
18	stdDBT1 vs stdDBT2	0.974	0.864	2.264	10.188	0.896	0.092	0.761	0.483	0.847
19	stdDBT1 vs stdDBT3	0.859	0.012	0.955	27.799	0.047	0.327	0.568	2.834	0.092
20	stdDBT1 vs stdDBT4	0.802	0.002	1.771	13.544	0.699	0.587	0.444	4.983	0.023
21	stdDBT2 vs stdDBT3	0.889	0.037	0.871	30.576	0.022	3.129	0.077	2.593	0.111
22	stdDBT2 vs stdDBT4	0.839	0.006	2.112	9.941	0.906	3.407	0.065	4.143	0.037
23	stdDBT3 vs stdDBT4	0.955	0.514	2.063	14.226	0.651	6.915	0.009	2.292	0.143

Table A. 34. The relationships between each breathing metric and its respective metric from every task (1-23 combinations).

	Variable Pair	S-W Statistic	S-W p-Value	Durbin-Watson	Ljung-Box Statistic	Ljung-Box P-value	Breusch-Pagan Statistic	Breusch-Pagan P-value	Rainbow Statistic	Rainbow P-value
24	stdDBP1 vs stdDBP2	0.843	0.006	1.541	29.506	0.03	0.133	0.716	5.500	0.018
25	stdDBP1 vs stdDBP3	0.869	0.017	1.488	16.376	0.497	1.781	0.182	2.980	0.082
26	stdDBP1 vs stdDBP4	0.924	0.153	1.625	9.585	0.92	3.143	0.076	1.071	0.475
27	stdDBP2 vs stdDBP3	0.961	0.622	1.638	21.44	0.207	3.273	0.070	1.305	0.371
28	stdDBP2 vs stdDBP4	0.961	0.612	2.053	13.364	0.711	4.027	0.045	6.276	0.012
29	stdDBP3 vs stdDBP4	0.870	0.018	2.403	9.081	0.938	0.129	0.719	0.531	0.815
30	mPH1 vs mPH2	0.938	0.272	Nan*	Nan*	Nan*	Nan*	Nan*	Nan*	Nan*
31	mPH1 vs mPH3	0.966	0.719	Nan*	Nan*	Nan*	Nan*	Nan*	Nan*	Nan*
32	mPH1 vs mPH4	0.955	0.517	Nan*	Nan*	Nan*	Nan*	Nan*	Nan*	Nan*
33	mPH2 vs mPH3	0.947	0.384	0.697	26.458	0.067	0.430	0.512	2.314	0.141
34	mPH2 vs mPH4	0.976	0.893	1.635	25.752	0.079	0.293	0.588	1.138	0.442
35	mPH3 vs mPH4	0.954	0.494	0.709	39.003	0.002	0.096	0.757	3.728	0.048
36	mPS1 vs mPS2	0.978	0.927	1.901	8.533	0.954	0.215	0.643	0.398	0.900
37	mPS1 vs mPS3	0.934	0.226	1.514	16.048	0.52	0.552	0.458	0.651	0.732
38	mPS1 vs mPS4	0.965	0.696	1.746	9.67	0.917	3.113	0.078	0.320	0.943
39	mPS2 vs mPS3	0.966	0.720	2.528	17.592	0.415	0.254	0.614	2.097	0.171
40	mPS2 vs mPS4	0.966	0.720	1.843	9.432	0.926	1.086	0.297	0.158	0.993
41	mPS3 vs mPS4	0.968	0.764	2.661	10.986	0.857	0.103	0.749	0.405	0.897
42	mTS1 vs mTS2	0.888	0.036	1.946	14.109	0.659	1.036	0.309	0.136	0.996
43	mTS1 vs mTS3	0.928	0.182	1.746	16.773	0.47	0.248	0.618	2.192	0.157
44	mTS1 vs mTS4	0.947	0.379	1.820	16.56	0.485	9.508	0.002	0.169	0.992
45	mTS2 vs mTS3	0.813	0.002	2.225	8.896	0.943	0.526	0.468	0.308	0.948
46	mTS2 vs mTS4	0.963	0.663	1.877	9.078	0.938	8.423	0.004	0.239	0.975
47	mTS3 vs mTS4	0.910	0.080	1.633	10.765	0.869	12.223	0.000	0.136	0.996

Table A. 35. The relationships between each breathing metric and its respective metric from every task (1-23 combinations).

*The Nan or not a number is a fault most likely cuased by a too high of a collinearit

Ingi Páll Eiríksson
8792283
Department of Information and Computing Science
Human-computer Interaction Master Thesis