universität wien

MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„How Are You Feeling? Influence of Interactive Emotional Cues on Empathic Experience in Computer-mediated Communication“

verfasst von / submitted by

Dinara Talypova

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna, 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 013
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Joint-Degree Middle European Interdisciplinary Master Programme in Cognitive Science UG2002
Betreut von / Supervisor:	doc. RNDr. Martin Takáč, PhD.

"Feelings are not substances to be discovered in our blood but social practices organized by stories that we both enact and tell."

— Michelle Rosaldo

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to doc. RNDr. Martin Takáč, PhD. for the supervision of this thesis and, especially, personal mentoring, fruitful discussions, and unlimited inspiration. I do believe that non-indifferent cooperation can create miracles – our working environment was the best ground not only for my professional but also personal growth.

I would like to thank Ouassim Fari, MSc – my kind and talented friend – for developing the app for this study, and his patient advice during the whole process. The thesis, as it is now, would not have been possible without his amazing help and patronage.

Finally, my deep gratitude goes to my parents – Indiana Talypova & Rishat Talypov – and my partner – Dmitrii Voinov – for their unfailing encouragement and faith in me. I especially appreciate this support – for the experienced professional, who leaves a lot aside, it is very scary to change. Taking the step to get into Mei:CogSci Master Programme was only possible because of your precious daily backstop and inspiration. I love you.

TABLE OF CONTENTS

1 Introduction 1

1.1 Technology for Communication 1

1.2 Emotions in HCI & CMC 2

1.3 Hypothesis, Research Question and Methods 4

1.4 Aim and Potential Impact 5

2 Literature Review and Concept Introduction 6

2.1 Interaction Process & Emotions 6

2.1.1 Social Interaction 6

2.1.2 Affective Interaction 8

2.2 Emotion Model & Conceptual Spaces Framework 11

2.2.1 Gärdenfors' Theory for Representing Information 11

2.2.2 Russel's Circumplex Model of Affect 13

2.3 Empathy 14

2.3.1 Origin(s) and Definition(s) of Empathy 14

2.3.2 Perceiving Empathy: Theory-theory vs. Simulation Theory 15

2.3.3 Empathy as Self-knowledge 16

2.3.4 Empathic Experience 16

3 Research Methodology 17

3.1 Short Overview of the Method 17

3.2 Preparational Procedures 19

3.2.1 Emotional map: Creating Affective Picture database 19

3.2.2 Developing the App 27

3.3 Experiment 31

3.3.1 Outline of the Experiment 31

3.3.2 Materials & Methods 33

3.3.3 Results 44

4 Discussion 59

4.1 Main Findings and Insights 59

4.2 Limitations and Suggestions for Improvement 61

4.3 Potential Impact and Future Research 62

4.4 Summary and Conclusion 63

5 References 65

6 Appendices 71

6.1 Abstract (English) 71

6.2 Abstrakt (Deutsch) 73

LIST OF FIGURES

Figure 1: App popularity according to iOS App Store rankings on March 16-18. Source: Apptopia. Credit: The NY Times, The Virus Changed the Way We Internet....2

Figure 2: Representation of an autopoietic system from "The Tree of Knowledge" by Humberto R. Maturana and Francisco J. Varela (left) and its adaptation with a focus on social interaction by Hanne De Jaegher from ICPS 2015 (right)....................................8

Figure 3: A typical example of a conceptual space by Gärdenfors to represent 'apple' concept (according to Banaee & Loutfi, 2014)...12

Figure 4: Eight affect concepts in a circular order (Russell, 1980, p. 1164) (left), and two-dimensional model of valence and arousal adapted from Russell (Walla et al., 2015, p. 408) (right). 13

Figure 5: The layout of the app that was developed to study interpersonal empathic communication with auxiliary interactive cues. Dragging the blue point changes the weather image of the agent (the bottom picture)..18

Figure 6: Four out of 55 pictures created for the research study to set up "Emotional map". 20

Figure 7: Socio-demographic data of respondents of "Emotional map" survey. 21

Figure 8: Univariate distributions of imagewise mean valence and arousal ratings (top row) and distributions of imagewise valence and arousal standard deviations (bottom row). 23

Figure 9: Q-Q plots of imagewise mean valence and arousal ratings with a 45-degree reference line. They show light tailed normal distribution of mean image ratings........24

Figure 10: Images with the highest and the lowest valence rating (top row), and the highest and the lowest arousal rating (bottom row). 25

Figure 11: Images with the lowest and the highest SD's for valence rating (top row), and the lowest and the highest SD's for arousal rating (bottom row). 25

Figure 12: Emotional map: image ratings in circumplex space, with valence (measured on a 1–7 Likert scale) on the x-axis and arousal (also measured on a 1–7 Likert scale) on the y-axis. 27

Figure 13: Voronoi tessellation for world airports map (left) and it's zoomed in version (right): each region (dyed in different colors) is closer to a particular airport than any other. Source: www.jasondavies.com/maps/voronoi/airports.......................................29

Figure 14: App matrix: original images on the emotional map augmented by newly created pictures....30

Figure 15: Socio-demographic data of respondents of the main study. 35

Figure 16: Data collection part of "Empathic Interaction Research Study" Google Form. 39

Figure 17: Guidance part of "Empathic Interaction Research Study" Google Form. 39

Figure 18: Partners' mood paths at the beginning and the end of conversations in the test group. Arrows show the direction of the participants' emotional states. Each color signifies the path of the partners in one pair. 46

Figure 19: Partners' mood paths at the beginning and the end of conversations in the control group. Arrows show the direction of the participants' emotional states. Each color signifies the path of the partners in one pair. 46

Figure 20: Approaching strategy paths for the test group (left) and the control group (right). The system of coordinates was compressed by width for better visibility. 47

Figure 21: "Catch-up" game strategy paths for the test group (left) and the control group (right). The agent approaches the initial emotional state of the partner, while the partner "goes away" in the other direction (not in the direction of the agent). The system of coordinates was compressed by width for better visibility. 47

Figure 22: Different directions paths in the test group (left) and the control group (right). The system of coordinates was compressed by width for better visibility. 47

Figure 23: Partners' mood paths (in 4 out of 14 pairs) that tend to synchronously approach the top-right corner (positive-excitement quadrant) in the test group. 48

Figure 24: Violin plots showing distribution of participants' responses for the questions of connectedness (top) and understanding partner's emotions (bottom) per group and partner type. Color dots are responses, black dots are means. 52

Figure 25: Test group. Juxtaposing the participants' "mood paths" – initial (darker) and final (brighter) points of coordinates connected by arrows – with the reposted subjective emotional states. Purple-to-pink paths - partners A, and orange-to-yellow paths - partners B. Inconsistent or illogical reports were taken into red frames. 54

Figure 26: Control group. Juxtaposing the participants' "mood paths" – initial (darker) and final (brighter) points of coordinates connected by arrows – with the reposted subjective emotional states. Purple-to-pink paths - partners A, and orange-to-yellow paths - partners B. Inconsistent or illogical reports were taken into red frames. 55

List of Tables

Table 1: Mean results in the distance of coordinates for test and control groups. Coordinates for the analysis were represented in decimal numbers. 44

Table 2: Mean results in changing the arousal level for test and control groups. 48

Table 3: Mean results in changing the arousal level for test and control groups divided by partner type. 49

Table 4: Mean results in changing the valence level for test and control groups divided by partner type. 49

Table 5: Mean results for questionnaire responses for test and control group with....... 51

Table 6: Mean results for questionnaire responses per group and partner. T-test and Wilcoxon test are applied for comparison within partner types (A and B), not within groups. 52

List of Abbreviations and Acronyms

CMC – Computer-Mediated Communication(s)

HCI – Human-Computer Interaction

HF – Human Factors

1 INTRODUCTION

1.1 TECHNOLOGY FOR COMMUNICATION

Technology communications are involved in many significant aspects of our life, and these communications include not only interaction with computers (Human-Computer Interaction, HCI), but also interpersonal interaction through computers (Computer-Mediated Communication, CMC). According to *Our World in Data*[1], the percentage of time we cooperate with others through technologies increases each year (*Broadband Subscriptions per 100 People*; *Daily Hours Spent with Digital Media*; *Mobile Cellular Subscriptions by Country*; *Percentage of Individuals Using the Internet by Country*; *The Rise of Social Media*). In 2020, due to the pandemic of COVID-19, we were observing a dramatic rise in consumption of social media and communication-related services that aimed to help people overcome the hardship of social distancing (Koeze & Popper, 2020). Figure 1 from NY Times represents this rise. Yet, studies showed we still were feeling lonely and isolated (Koeze & Popper, 2020; Ortiz-Ospina & Roser, 2020). What special do we have in physical real-life communication that we cannot transfer into the digital world? Why does it still not feel "the same" even though the quality of signal may be high? Some scientists believe that the reason is in limited amount of senses we use in computer-mediated communication (nowadays, we mostly interact with auditory and visual modalities) and try to experiment with adding tactile/sense of touch (Bickmore et al., 2010; Bruns Alonso et al., 2013), or even smell (Cozza et al., 2011; Gosain & Sajwan, 2014). Other researchers suggest enriching technology-mediated communication by providing extra contextual information (Guo et al., 2019; Kim et al., 2014). There are also supporters of the idea of simulating "ephemerality" of human-to-human communication, that is to dissipate the messages (i.e., do not keep track of the conversation history) as soon as they were sent - just like in real-life conversation (Xu, 2017). Even though the proposed assumptions have plausible premises and interesting results, they so far go with the idea of replicating real physical communication. The idea of this thesis, on the contrary, is to try enriching communications (that is, to increase

---

[1] *Our World in Data* and the *SDG-Tracker* are collaborative efforts between researchers at the University of Oxford (Oxford Martin Programme on Global Development), who are the scientific editors of the website content; and the non-profit organization Global Change Data Lab, who publishes and maintains the website and the data tools.

empathic experience and/or connectedness) by using the advantages of new technologies. So, to employ the benefits of a programmed world that real-life interaction cannot provide. However, the benefits of the tool (in this case, computer) are still based on the general understanding of how communication works.

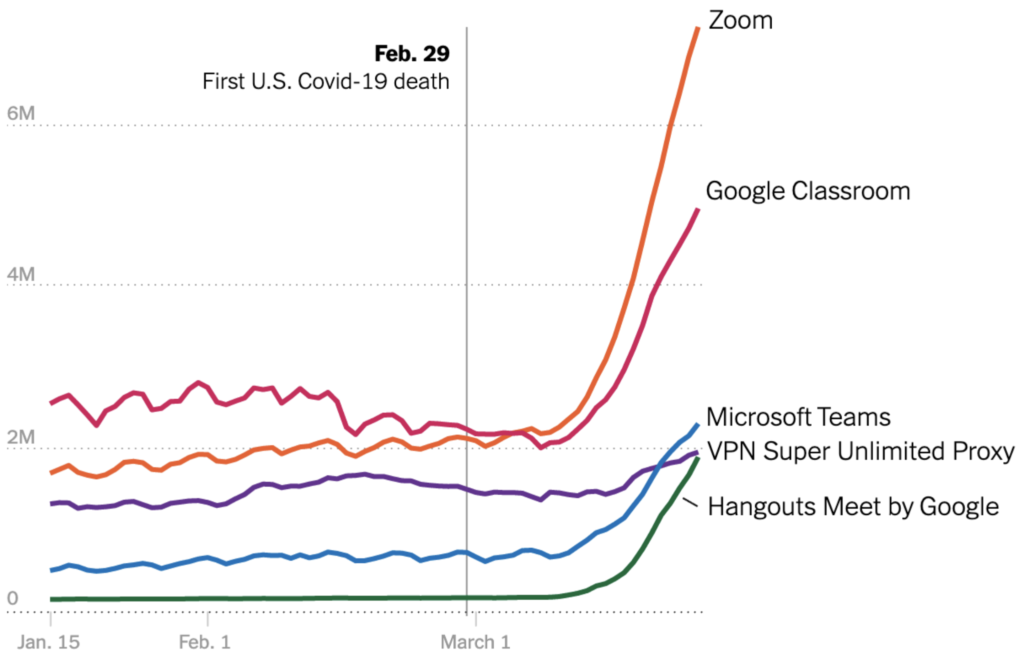

Figure 1: App popularity according to iOS App Store rankings on March 16-18. Source: Apptopia. Credit: [The NY Times, The Virus Changed the Way We Internet]

1.2 Emotions in HCI & CMC

One of the important aspects of communication involves emotions. In the past century, emotions were often considered a "distractive" aspect for human cognition and tended to be hidden in the shadows of consciousness, rational decision-making, and understanding others' intentions. Human beings were wished to be seen as rational agents - and the possibility to reason "objectively" was considered as our main advantage in the animal world. Hence, emotions were either neglected (as not worth studying) or considered as a reason for cognitive biases and impulsive solutions (so, some "antidote" should be discovered). Thus, the first communications through the devices were mostly functional (Jeon, 2017; R.W. Picard et al., 2001; Soegaard & Dam, 2013): the computer should serve a particular goal and be understandable for the user.

However, at the end of the 20th century a new wave of interest in the role of emotions rose. For instance, the works of neurologist A. Damasio (Damasio, 1996) showed that emotional processes were highly involved in reasoning and decision making. Moreover, empathy - "the capacity to understand and enter into another person's feelings and emotions or to experience something from the other person's point of view" (Colman, 2015, p. 245) - has been seen as an important ability to gain a grasp of the content of other people's minds and to predict and explain what they will think, feel, and do (Maibom, 2017). All these findings forced researchers to look "beyond the cognitive" and to understand affective aspects of communication experience.

The first voices for including emotional aspects in HCI & CMC sounded in relation to aesthetics, visual attractiveness, and consumer's feeling. The concepts of Emotional Design (Norman, 2004), Kansei Engineering (Nagamachi, 1995), Sensibility Ergonomics (Lee et al., 1999), Hedonomics - "pleasure laws" - (Hancock et al., 2005; Helander et al., 2003), and Affective Computing (Rosalind W. Picard, 1997) were introduced by designers, Human Factors engineers, and computer scientists. The latter approach, first proposed by Rosalind Picard in 1997, aims to study and develop devices that can recognize, interpret, process, and simulate human affects (Rosalind W. Picard, 1997). The prospective machine should interact with the user and have the ability to simulate empathy. However, before programming and implementing these "emotional intelligent" machines researchers first have to understand the origin of affects. The used methods for researching and measuring emotions in Affective Computing are broad and innovative: computer vision (e.g., for facial recognition), EEG, body temperature, cortisol and heart rate measurement, and many more (Jeon, 2017, p. 19). Scientists try to fixate human affects in different types of environment, put the labels, and teach machines to interpret them. However, these methods are plausible if the researcher works with emotions as static (not dynamic), objective (individual differences in reaction are negligible), and existed in a vacuum (not related to the environment/broader context).

The alternative view on the study of affect draws upon phenomenology and treats emotions as a dynamic interaction process. This approach was manifested by Kirsten Boehner et al. in "How emotion is made and measured" (Boehner et al., 2007) and introduced under the term "Affective Interaction". According to the researcher, "instead of designing systems to decode and encode affective signals, systems are set up to engage users in interactions and help them to understand, reflect on, and experience their emotions in new ways." (Boehner et al., 2007, p. 284). Dynamic interactive view on emotions intersects with the works of Hanne De Jaegher who studies the relation of social cognition and interaction. Jaeger argues that social interaction plays an enabling or constitutive role in social cognition. The idea is that mental processes cannot be fully grasped in a vacuum, and "meaning is generated and transformed in the interplay between the unfolding interaction process and the individuals engaged in it." (De Jaegher & Di Paolo, 2007, p. 485). This thesis is based on the same assumptions about emotional experience.

1.3 Hypothesis, Research Question and Methods

Taking into account the above statements, I designed an experiment that aimed to measure empathic communication between people with and without extra interactional aspect (here: cues). Communication was mediated by technologies: it took part by online video-conference tool and involved a mobile app that was developed for this study. In a nutshell, the analytical objective of the thesis is to explore whether additional interactive cues enhance the empathic experience in computer-mediated communication.

The hypothesis of the study is that partners who are communicating in an enriched interactive environment should have a better mutual emotional understanding. For the experiment, two groups of partners (test and control) were asked to communicate through audio headset. In the test group, both partners used a display of a smartphone (the app) to send interactive signals (weather pictures) while having an audio conversation. Each picture reflected an “emotional state” of a partner in real-time. In the control group, partners did not exchange pictures in real-time, however they still used the app for themselves to fixate their "emotional state"[2]. For measuring the difference in two groups, I used quantitative methods: Likert-scale survey about partners' experience, and statistical analysis of "emotional closeness" to each other based on the theory of Conceptual spaces[3] (Gärdenfors, 1996). Additionally, I used some qualitative methods (self-reflection, joint reflection, interpersonal communication) to review the subjective experience of the participants.

---

[2] More information about the design of the experiment can be found in Research Methodology chapter.

[3] "Emotional closeness" is measured thanks to the developed software. For further details see Outline of the Experiment.

1.4 Aim and Potential Impact

In this research, I abandon the idea to use machine as a final receiver and interpreter of the emotional processes not only because it takes much more complex methods (e.g., bearing in mind individual characteristics, environment, and (inter)actions), but also because the main purpose of using technologies in this approach is to help people to evoke and process empathy through the channel they tend to use more often.

Besides helping to improve digital empathic experience - oftentimes impoverished - this research may benefit general understanding of emotions and empathy. Moreover, interpersonal communication through technologies help us to overcome some experimental challenges, e.g., better controlled conditions like amount of modalities, or regulated interactive cues, or smoother (less distracting) measuring process (created software works as a measurement tool while being a communication tool).

2 LITERATURE REVIEW AND CONCEPT INTRODUCTION

2.1 INTERACTION PROCESS & EMOTIONS

2.1.1 SOCIAL INTERACTION

In this section, I would like to introduce the views on interaction processes by a philosopher of mind and of cognitive science Hanne De Jaegher. In her and her colleagues' work from 2010, she notices an important tendency in social cognition research, stating that "the focus goes away from the individual mind and toward embodied and participatory aspects of social understanding" (De Jaegher et al., 2010, p. 441). The researcher describes her vision of social cognition and that it expands from just explaining the other's actions: "It involves understanding others but also understanding with others" (De Jaegher et al., 2010, p. 442). Noticeably, by "understanding" the philosopher means the ability to act appropriately in specific circumstances, and this process may not involve verbal identification of the intentions. De Jaegher defines social cognition as:

”*General term used to describe different forms of cognition about, or actions in regard to, agents or groups of agents, their intentions, emotions, actions and so on, particularly in terms of their relation to other agents and the self.*"

Hanne De Jaegher argues that interactive processes that occur in our communication are more than a context for social cognition: they can complement and even replace individual mechanisms. She quotes several works where the importance of interactive processes has been highlighted: the studies of different forms of coordination in dynamical systems approaches (Richardson et al., 2007; Shockley et al., 2009) and developmental studies (Fogel & Garvey, 2007; Trevarthen & Aitken, 2001) signifying that coordinated behavior occurs from the mutual regulation of a social experience. Even mirror neurons could function differently in various interactive contexts: depending on whether an interactive situation presents conflict or not (Naotaka et al., 2007). De Jaegher makes a conclusion that mirror neurons "could develop as a result of the agent’s skilful involvement in social interaction rather than being the wellspring of capacities for social understanding" (2010, p. 445), in other words, mirror neurons are the product of active social interaction and not the static embedded "tool" of social cognition. The philosopher claims that social interaction plays an enabling or constitutive role in social cognition. She defines social interaction "as a co-regulated coupling between at least two autonomous agents, where: (i) the co-regulation and the coupling mutually affect each other, constituting an autonomous self-sustaining organization in the domain of relational dynamics and (ii) the autonomy of the agents involved is not destroyed (although its scope can be augmented or reduced)" (2010, pp. 442–443).

These views are closely connected with the enactive approach to cognition introduced by Francisco Varela, Evan Thompson, and Eleanor Rosch (Varela et al., 1991). In enactivism, cognition is constituted from active interaction (coupling) between an organism and the environment. It rejects the classical passive input-output scheme of cognitivism where cognition is explained as an internal symbolic representation of the external world. In enactivism, the organism is a center of activity.

Self-individuating – distinguishing oneself from the environment and others – is a key attribute of an enactive agent, and it is generated and maintained by interactive *autonomy* – the way how the system adapts and self-sustains. Maturana and Varela define an *autopoietic system* that "generates and specifies its own organization through its operation as a system of production of its own components" (Varela et al., 1974, pp. 187–196). *Adaptivity* is the property of maintaining itself "through constant structural and functional change" (Shapiro, 2014, p. 68). To adapt, the system has to establish some requirements or norms that have to be satisfied to maintain itself stable. It creates and sustains proper precarious conditions for the system, and thus realizes *internal equilibrium*. The system has to define meaning and significance in the organism-environment relationship through *sense-making*.

In enactivism, it is proposed that perception experience is constituted by the actions of a situated cognitive agent. In other words, we portrait the world by virtue of mutual interactions between the physiology of the organism, its sensorimotor circuit, and the environment.

The representation of the autopoietic system from "The Tree of Knowledge" (Maturana & Varela, 2008) is depicted in Figure 2 (left).

Hanne De Jaegher develops the views from an enactive approach further and adapts the autopoietic system by Maturana & Varela from a social interaction point of view

(Figure 2 (right), (De Jaegher, 2015): this is not just an autonomous system that maintains the equilibrium and constantly interacts with the environment, this is a coupling between at least two autonomous agents that co-regulate each other while staying autonomous, and the world that is also involved in the interactive process (see De Jaegher definition of social interaction in the beginning).

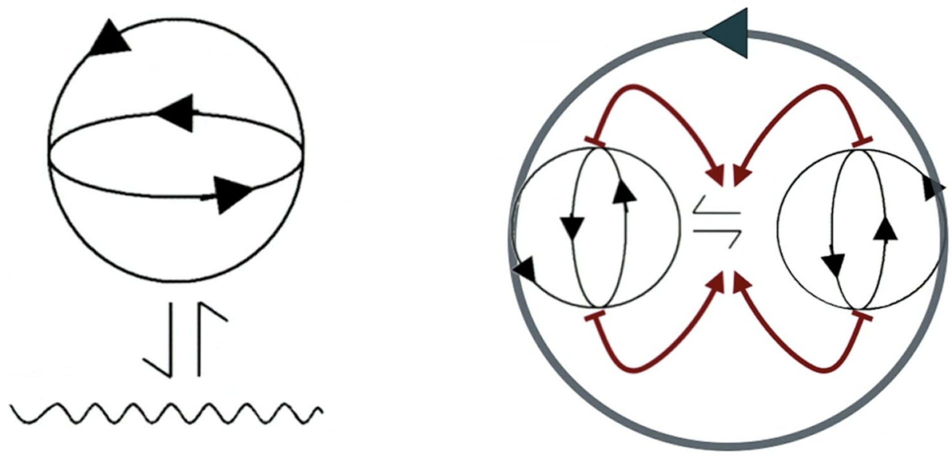

Figure 2: Representation of an autopoietic system from "The Tree of Knowledge" by Humberto R. Maturana and Francisco J. Varela (left) and its adaptation with a focus on social interaction by Hanne De Jaegher from ICPS 2015 (right).

Social interaction is a sophisticated process that involves varying sets of circumstances, behavioral strategies, individual differences, various timescales, etc. Essential to interaction is that it involves engagement between agents. Engagement helps to capture the qualitative aspect of social interaction. And, according to De Jaegher, it corresponds to "fluctuating feelings of connectedness with the other" (2010, p. 442). Social interaction theory plays a core role in my thesis and in the design of this research. I presume that an enriched interactive environment will help enhance engagement and connectedness between participants in my experiment.

2.1.2 AFFECTIVE INTERACTION

Eva Hudlicka in her chapter "Computational Modeling of Cognition–Emotion Interactions" (Jeon, 2017, pp. 383–435) stresses out that it is very difficult to propose one definition for emotions. However, most emotion researchers would accept a working definition that views emotions "*as states that reflect evaluative judgments of the environment, the self and other agents, in light of the agent's goals and beliefs, and which motivate and coordinate adaptive behavior*" (Jeon, 2017, p. 385). Here, she states, the terms “goals” and “beliefs” are used in a generic sense: goals reflecting desirable states, and beliefs reflecting current knowledge. Moreover, emotion research includes several relevant constructs, such as feelings, moods, emotions, and affect. These notions are distinct, but in this study, they are going to be used interchangeably, as a frequent practice in the field (Jeon, 2017, p. 4).

For a long time, emotions and affect were not the dominant topics for research in human factors (HF) and the human-computer interaction (HCI) fields (Jeon, 2017, p. 3). The primary paradigm was the information processing approach of traditional Cognitive science, where there was no place for emotions (Jeon, 2017, p. 3). However, over the last two decades, significant activity developed in emotion research. New studies showed that it is impossible for people to assess the event or make actions without engaging their emotional system (Nass et al., 2005) and, moreover, emotion-related brain disorders can lead to inability of making decisions (Damasio, 1996). Currently, emotional understanding and empathy are one of the main topics in affective computing research (e.g., (El Kaliouby, 2005). Application domains of affective computing include computer-mediated education, computer-assisted learning, entertainment, culture and art, health, mental issues (e.g., autism). In the business field though is often used for analyzing pleasurable and displeasurable experiences of the products or services.

The shifted focus in Cognitive science from the representational point of view to the 4E approaches extended HF and HCI agendas and began to move the field from an informational to an interactional model. Kirsten Boehner in her work "How emotion is made and measured" (2007) argues that the interactional approach that started to prevail in HCI, refreshed the field in three ways. First, it helps to see emotions "as culturally grounded, dynamically experienced, and to some degree constructed in action and interaction" (2007, p. 276). Second, as an interface paradigm, it shifts the focus from understanding humans' emotions by computer to helping users to better process and experience their own emotions. Therefore, third, it changes the evaluation criteria for the technology: "Measures of success for such systems therefore do not focus on whether the systems themselves deduce the ‘right’ emotion but whether the systems encourage awareness of and reflection on emotions in users individually and collectively." (2007, p. 276)

One of the examples of affective interaction studies is the work from Petra Sundström and her colleagues where they tested the eMoto concept (Sundström et al., 2005). eMoto is an extended SMS-service for the mobile phone that lets users send text messages, but the text is also enriched with a specific background. This background is colored depending on the emotion that the sender tries to express. To manage the color and the shape of the effect, the user uses gestures. The study is relatively old for technological research (in 2021, most users communicate through smartphones, where they can send smileys and GIFs), however, it may be considered as brand new for interactive emotion research. Moreover, an important aspect of this study compared to smileys in messages is the possibility of the ambiguous, open-ended interpretation of the sender's emotions, and reflecting on the receiver's perceptions.

Since new success criteria and goal for emotion researchers is to make emotional experiences available for user's reflection, and not just educate machines to interpret them, I propose that the affective interaction approach will help in this task. Working with emotions as a dynamic and socially constructed process is one of the main premises of this thesis.

2.2 EMOTION MODEL & CONCEPTUAL SPACES FRAMEWORK

2.2.1 GÄRDENFORS' THEORY FOR REPRESENTING INFORMATION

In the book "Conceptual Spaces: The Geometry of Thought" (2000) Peter Gärdenfors describes his theory of conceptual spaces that aims to explain how cognitive systems represent meaning. In a nutshell, the theory is alleged to avoid the problematic aspects of the symbolism (i.e., a cognitive system as a Turing machine) and the associationism (i.e., a cognitive system that models associations among different kinds of information elements) and to work as a bridge between these two approaches.

According to Gärdenfors, a conceptual space is based on a number of quality dimensions, e.g., brightness, temperature, length, taste, etc. It is built up from geometrical structures rather than symbols or associations. This means that similar objects are "situated" closer to each other in n-dimensional space. And properties of objects are represented as particular values along dimensions. To illustrate this idea Gärdenfors uses the example of color. The cognitive representation of color can be described by three dimensions: hue, saturation, and brightness. While the last two have a linear structure (from grey to intensive color and from white to black), the first one (hue) is a polar coordinate where each direction from the center of the color circle expresses a certain color tone.

The main function of the dimensions is to represent objects' qualities. Dimensions can be integral (i.e., when some value for an object is assigned on one of the dimensions, it must be also assigned a value on the others) or separable (i.e., when dimensions are independent of one another). For example, saturation and hue are integral, while form and hue are separable. Based on this distinction, Gärdenfors defines a domain as "a set of integral dimensions that are separable from all other dimensions" (Gärdenfors, 2011, p. 2). Hereby, a conceptual space can be described as "a collection of quality dimensions divided into domains" (Gärdenfors, 2011, p. 2).

However, as the author of the theory notes, the dimensions of a conceptual space should not be treated as fully independent entities and may co-vary. For instance, in the fruit domain, taste and color are seen as integral. In Figure 3 there is a neat illustration of Gärdenfors' framework based on the apple example (Banaee & Loutfi, 2014, p. 12).

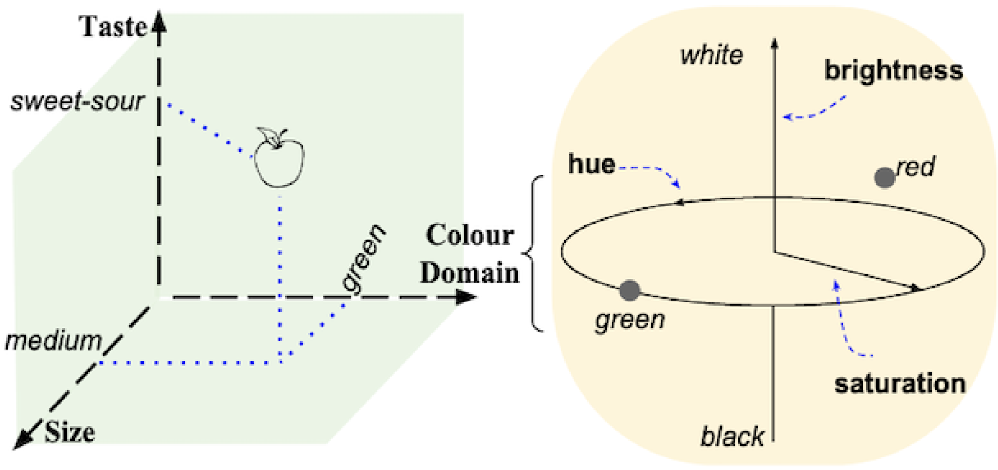

Figure 3: A typical example of a conceptual space by Gärdenfors to represent 'apple' concept (according to Banaee & Loutfi, 2014).

In Gärdenfors theory of conceptual spaces, he emphasizes a spatial structure to the properties of objects. That is, if two objects A and B are instances of some concept, then any object placed between A and B will also be an instance of the same concept.

Substantially, the dimensions are understood to be cognitive and infra-linguistic, which means that one can represent the qualities of objects "without presuming an internal language in which these qualities are expressed" (Gärdenfors, 1996, p. 6). In other words, one can build a concept for an object without knowing the terminology for it. Particularly, the infra-linguistic statement is important for this thesis.

In this thesis, I use the theory of conceptual spaces to build an "Emotional map" for the experiment software[4]. Even though Gärdenfors does not linger on emotional aspects with his framework, he mentions it as a working domain to the theory (Gärdenfors, 2000, p. 162, 2011, pp. 6–7). At the preface of his book (2000, p. IX), the author states that his views are based on four disciplines: philosophy, computer science, linguistics, and psychology. The theory of conceptual spaces is strongly associated with psychological discipline and, in particular, with space models of emotions. One of the pioneers in this approach was psychologist James A. Russell.

[4] See Emotional map: Creating Affective Picture database section for details.

2.2.2 <u>RUSSEL'S CIRCUMPLEX MODEL OF AFFECT</u>

In psychology, James Russell is known for his dimensional approach for emotions (Russell, 1980): he proposed a circumplex model of affect. According to the scientist, all emotions are "interrelated in a highly systematic fashion" (1980, p. 1161) and can be represented by a spatial model. The model is circumplex since Russell describes affective concepts using arc degree: pleasure (0°), excitement (45°), arousal (90°), distress (135°), displeasure (180°), depression (225°), sleepiness (270°), and relaxation (315°). Yet, it can be described as a two-dimensional system of coordinates where the horizontal dimension represents valence (from negative to positive) and the vertical dimension represents arousal (from low to high excitement). Figure 4 represents Russell's original eight affect concepts in a circular order (Russell, 1980, p. 1164) and a two-dimensional model of valence and arousal adapted from Russell (Walla et al., 2015, p. 408) - for illustrative purposes.

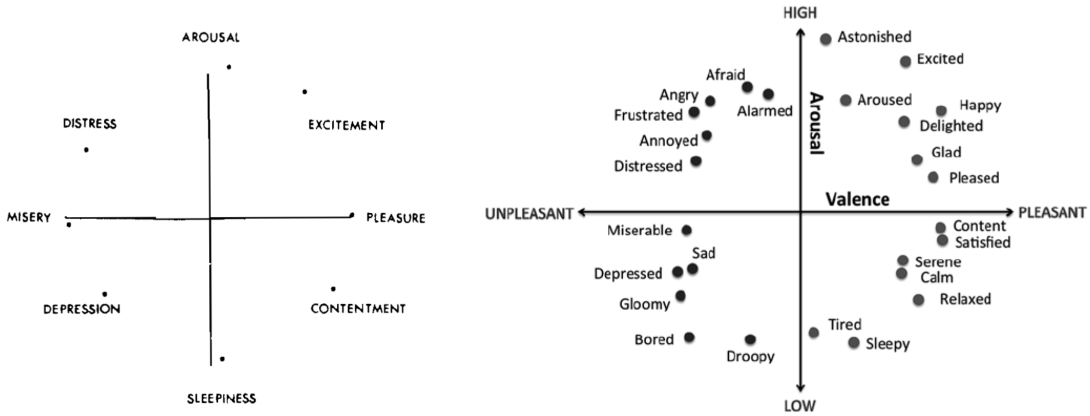

Figure 4: Eight affect concepts in a circular order (Russell, 1980, p. 1164) (left), and two-dimensional model of valence and arousal adapted from Russell (Walla et al., 2015, p. 408) (right).

Since the first Russell's proposal of a space model of affect in 1980, other dimensional models (Bradley et al., 1992; Mehrabian, 1980; Plutchik & Kellerman, 1989; Scherer, 2005; Watson D & Tellegen A, 1985) with more elaborate structure, shapes and amount of dimensions were introduced. However, I decided to stick with Russell's two-dimensional model in the research design. It helped to avoid overload in the experiment and to simplify the task for the participants. More details about the experiment design are provided in Preparational Procedures section.

2.3 EMPATHY

2.3.1 ORIGIN(S) AND DEFINITION(S) OF EMPATHY

The term "*empathy*" is defined very broadly in different scientific fields and directions, so when the researcher works with this concept, she has to explicitly state what it means in each case. Empathy has its origins from the concept of "*sympathy*" first introduced and described by David Hume in "A Treatise of Human Nature" (1739) where he tried to explain "a variety of psychological phenomena, including the transmission of emotion from one person to another" (317, 319, 363, 592). Hume believed that sympathy is a natural and automatic process "that enables the fast and instantaneous spread of emotion" (Coplan & Goldie, 2011, p. XI), which is close to the modern aspect of *mirroring*. Later, in "The Theory of Moral Sentiments" (1759), Adam Smith developed the concept of sympathy further. But, according to Smith, sympathy involves imaginative perspective-taking, i.e., the high-level experience of others' emotions. The English term "*empathy*" first appeared in Edward Titchener's "Elementary Psychology of Thought Processes" (1909) where he explained empathy as the process of not only seeing somebody's particular emotions but also to "feel or act them in the mind's muscle" (21).

Heidi L. Maibom in her book "The Routledge Handbook of Philosophy of Empathy" introduces more specific term for the purposes of this study - "*Affective empathy*". It is characterized by "a range of emotional responses we can have to what others feel or the situation they are in, which include sympathy, empathic anger, and contagious joy" (2017, p. 22). According to Maibom, this stands in contrast to 'cognitive empathy', which is typically used to describe our ability to grasp others' thoughts, feelings, or beliefs. Maibom (2017, p. 22) claimed the following statement about affective empathy:

Person S empathizes with person O's experience of emotion E in situation C if S feels E for O as a result of believing or perceiving that O feels E, or imagining being in C.

At any rate, there are plenty of discussions about what should be considered as empathy and hot debates regarding different theories, but the constraints of this thesis do not let us discuss them in detail. Still, I would like to briefly describe two mainstream points of view as the most contrasted one to another.

2.3.2 Perceiving Empathy: Theory-theory vs. Simulation Theory

Even though there is little consensus about what counts as empathy, most researchers unite in stating "some type of shared emotion" as an essential component (Coplan & Goldie, 2011, p. XXIII). However, the difference is in explaining how this shared emotion arises: from the early works, some associated it with the concept of *Verstehen* (understanding), especially in the phenomenological and hermeneutic traditions, while others saw it as the capacity to directly experience others' thoughts, emotions, and desires (Coplan & Goldie, 2011, p. XIII). These two original camps are associated with more general theories of making sense of the outside world: *theory-theory* and *simulation theory*. According to theory-theory, one creates a concept (a theory) regarding what the others do, feel or desire, and, thanks to this, explains the outside world. That is, the theory about other minds helps the person to understand others' emotions, but not obviously perceive these emotions by oneself. On the contrary, in simulation theory, one makes sense of others' behavior (and feelings) by engaging with others' minds, and 'simulating' these mental processes. That is, by imagining being in that other person's position, one can determine what the other will feel, think, and decide.

Another important aspect is the division into two kinds of empathy: *higher-level empathy* and *lower-level empathy*. Lower-level empathy is mostly understood as a biological, unconscious process. Other terms for lower-level empathy include *basic empathy* (Steuber, 2006), and *mirroring* (Goldman, 2011). Higher-level empathy is understood as a conscious process of "putting yourself in other's shoes". Other terms for higher-level empathy include *re-enactive empathy* (Steuber, 2006), *reconstructive empathy* (Goldman, 2011), and *perspective-shifting* (Goldie, 2011).

During the past two decades, some of the most important contributions to the study of empathy have come from the field of neuroscience (Coplan & Goldie, 2011, p. XXXVI). And one of the most revolutionizing discoveries was 'mirror neurons.' What makes these neurons special is that they are activated both when an individual performs a particular type of action (e.g., grasping an object) and when an individual observes another performing that type of action. Thus, we can say that we 'mirror' others at a neurological level. Moreover, later research revealed that there are more types of mirror neurons that are responsible not only for actions, but also for sounds, and even for emotions, pain, and touch (Coplan & Goldie, 2011, p. XXXVI).

It is considered that discovering mirror neurons won favor for the simulation theory. Although it should be said that it moved forward mostly for lower-level empathy, there are still some scientists who suggest that a sophisticated neural system may support complex forms of mindreading and empathizing, also including higher-level empathy (Coplan & Goldie, 2011, pp. 45–57).

In this research, I follow the simulation theory as the explanation of empathy. That is, emphasizing to somebody means to some extent experiencing somebody's emotions by oneself. This concept is crucial for choosing the research method and designing the experiment.

2.3.3 Empathy as Self-knowledge

As a side note, I want to introduce some thoughts of Edith Stein who argued that empathy "enables us to understand others but also to understand ourselves as others experience us" (Stein & Stein, 1989, p. 60). That is, besides understanding the emotions/beliefs/desires of others, we at the same time, gain self-knowledge by trying to know/perceive how others experience ourselves. This self-knowledge is vital for our development (Stueber, 2006; Zahavi, 2001).

I am not going to go deep into this aspect of empathy in my research, yet Stein's argument may be interesting for the discussion part of this thesis.

2.3.4 Empathic Experience

In this thesis I use the term "empathic experience" instead of "empathy". In the research, I try to understand whether some changes in the interactive environment help two people to better perceive each others' emotions in the particular situation and with the particular conversation. I do not try to enhance participants' general ability to empathize but make their communication more comfortable where they can feel closer to each other, gain richer emotional experience and feel stronger connectedness.

3 RESEARCH METHODOLOGY

3.1 SHORT OVERVIEW OF THE METHOD

In my research, I designed an experiment that aimed to measure empathic communication between people with and without extra interactional cues. Communication was mediated by technologies: it took place by online video-conference tool (where the only audio channel was used) and involved a mobile app that was developed for this study.

The app is web-based and represents a screen with two rectangle pictures of weather and a so-called valence-arousal scale with a point to drag. The valence-arousal scale is the two-dimensional system of coordinates that was introduced in chapter Russel's Circumplex Model of Affect. The horizontal dimension represents valence - it is changing from negative to positive from left to right side. The vertical dimension represents arousal - it is changing from a low to a high level of excitement from the bottom to the top of the scale. When the agent drags the point in the system of coordinates it changes one of the weather pictures - the picture of the agent. The second rectangle image is manipulated remotely by the agent's partner who uses the app in the same way. This means, two partners can constantly change the pictures in online-regime and the app will update it on both devices in real-time. It works smoothly so that partners could update their images often during the conversation. The position (coordinates) of each weather image is described in detail in the chapter Emotional map: Creating Affective Picture database below. Two layouts of the app are presented in Figure 5.

In the experiment, there was a test group and a control group. In both groups, two partners had to communicate with each other through an audio channel (about a predefined topic) for a limited amount of time. The test group used the app exactly how it is explained above. That is, the test group participants exchanged auxiliary interactive cues with each other. These cues were weather pictures corresponding to particular points of coordinates on the valence-arousal scale. While the control group partners did not exchange weather images in real-time, however, they still used the app for themselves to fixate their "emotional state". That is, they dragged the point and changed the pictures, but the partner's rectangle was not updated - the new point of coordinates was sent only to the server for future analysis. After the conversation, every participant had to fill in a short questionnaire regarding the quality of the conversation. There was also a group-reflection part with debriefing.

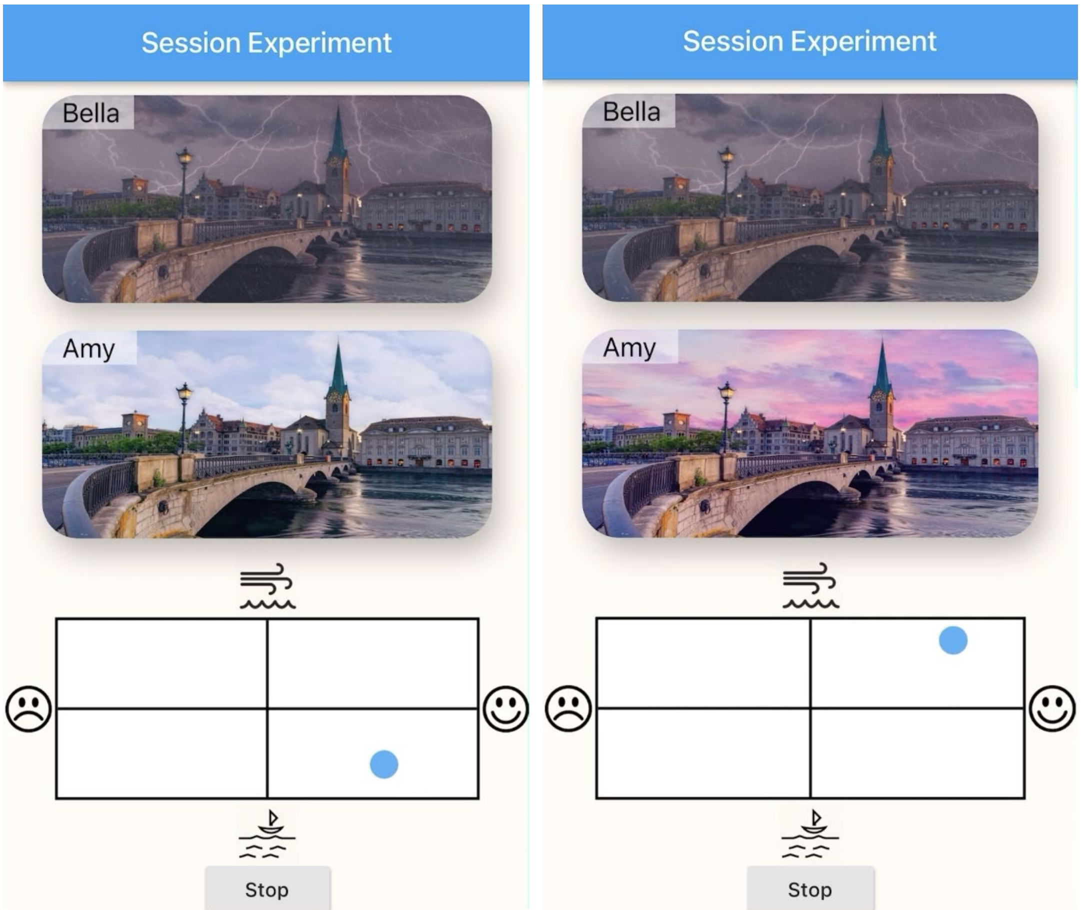

Figure 5: The layout of the app that was developed to study interpersonal empathic communication with auxiliary interactive cues. Dragging the blue point changes the weather image of the agent (the bottom picture).

3.2 Preparational Procedures

3.2.1 Emotional map: Creating Affective Picture database

In the introduction, I mentioned that each picture in the app reflected an "emotional state" of a partner. That was made possible thanks to the valence-arousal scale, which we will call "Emotional map". The images on this scale are situated based on the survey I have conducted in advance. The survey is a replication of research that was published by Kurdi et al. (2017). The study introduces the Open Affective Standardized Image Set (OASIS) - an open-access online stimulus set for 900 images that were rated on two affective dimensions (valence and arousal). According to the authors, the valence-arousal scale "covers much of the circumplex space and is highly reliable and consistent across gender groups" (Kurdi et al., 2017, p. 457), that is why this spatial model was used as the most relevant to the study.

- ***Materials***

In my research, I aimed to assign the coordinates for valence and arousal dimensions to the weather pictures, so that I could implement them later into the app. The respondents were asked to rate the image of a city landscape (Zurich) that was presented in 55 weather conditions. The original image was taken from a website for sharing photos and illustrations Pixabay (https://pixabay.com). It was placed under Pixabay License which means the picture can be used for free for commercial and noncommercial use without restrictions on modification (Pixabay License, n.d.). To modify the image and create 55 different weather conditions from it, I used the mobile app Lightleap (https://lightricks.com/products). It changes the sky and applies a special filter depending on a chosen part of the day and weather state. Several examples of created pictures are represented in Figure 6.

- ***Participants***

Respondents for this survey were recruited by social networks (random groups and communities) and word of mouth (acquaintances and other social circles of the researcher). The participation was totally voluntary and it helped to assure that the answers were honest and the respondents were sufficiently motivated to help scientific study. Altogether, 317 participants took part in the survey (151 participants rated arousal

Figure 6: Four out of 55 pictures created for the research study to set up "Emotional map".

dimension, and 166 participants rated valence). Some socio-demographic data about respondents were collected. It is worth mention that there are some shifts in the data: female respondents prevail (78.9% females vs. 20.2% males and 0.9% other answers), Russian is a native language for almost a third of participants (29.8%), and most of the respondents are from Europe (87.1% chose Europe as a current residence region and 79.2% stated that they have lived longest period in Europe). Age group and education level responses are distributed more or less uniformly with some dominance of 25-34 years old group (51.4%) and Master degree group (45.4%).

Socio-demographic data were gathered to identify possible factors that may influence picture assessment, e.g., linguistic specifics, gender, generation, or cultural differences. Nevertheless, the overview of the responses showed that participants rated the images similarly irrespective of most of socio-demographic parameters[5].

The distribution of socio-demographic responses is provided in Figure 7.

[5] I applied multiple regression models to predict the mean rating of each participant based on one's demographic data. Only a group of respondents under the age of 18 (3 participants) for arousal ratings (but not valence) showed significant result ($p < .000$): participants of this age category had increased mean result by 1.99 points. The adjusted R2 of the model was .308.

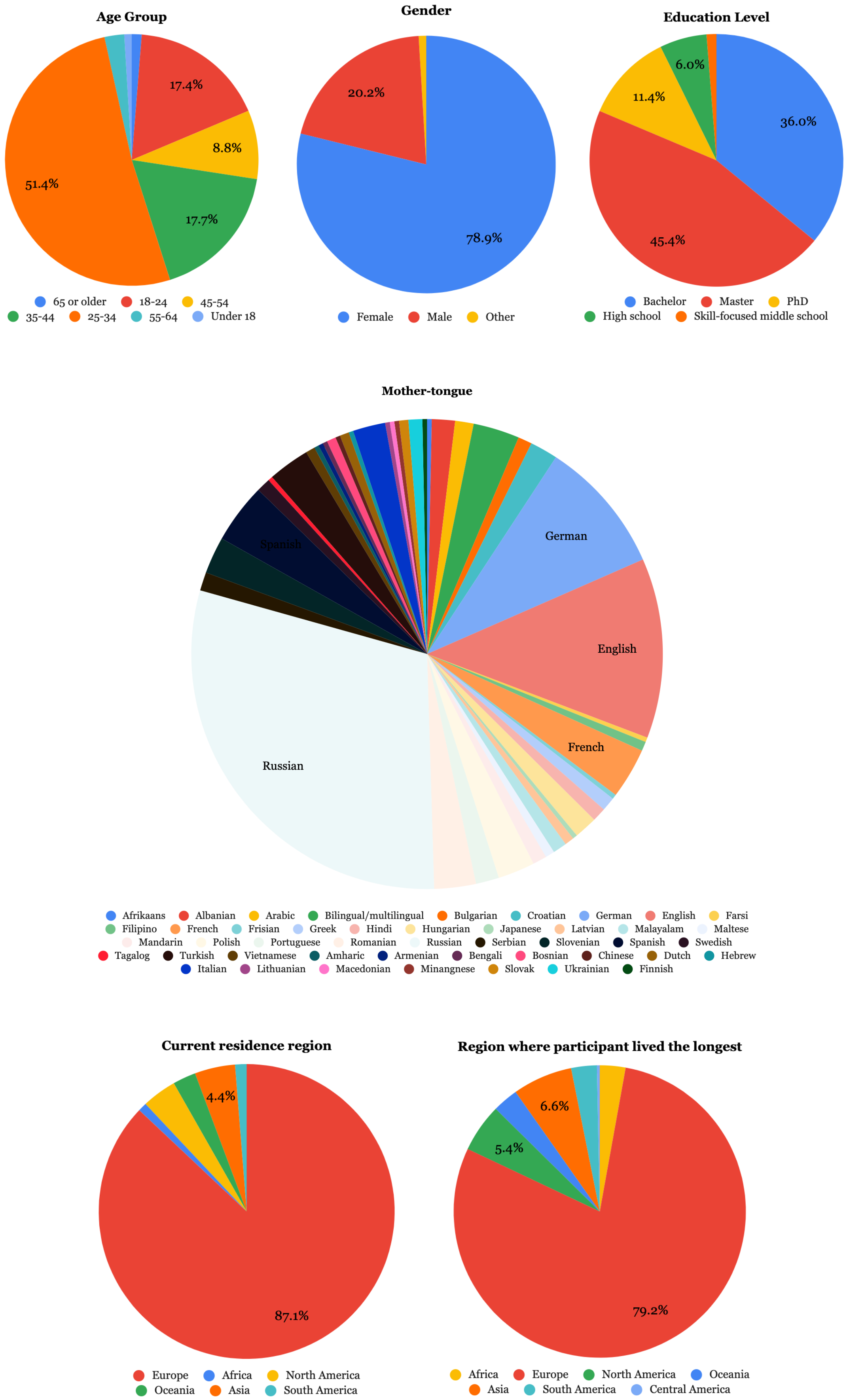

Figure 7: Socio-demographic data of respondents of "Emotional map" survey.

- ***Main study***

Participants were asked to rate each picture (same city view with different weather conditions) on only the valence (one subgroup, 166 randomly chosen participants) or only the arousal dimension (another subgroup, the remaining 151 participants). The separation of the two scales was done to avoid possible unconscious confusion between them. Participants received a general description of the study, a definition of the dimension with examples, and instructions. Instructions asked participants to indicate the level of valence or arousal intrinsic to each image (*What level of emotional intensity does this picture reflect?* or *What emotion does this picture reflect?*) using a 7-point Likert scale. Even though this research follows an infra-linguistic approach, I decided to use a verbal scale to help respondents navigate through the survey. According to Bradley & Lang (1994), verbal scales for valence and arousal are very highly correlated with Self-Assessment Manikin - a non-verbal pictorial assessment technique for the pleasure, arousal, and dominance - which is normally used for the sets of emotional stimuli, e.g., the International Affective Picture System (IAPS; Bradley & Lang, 1994, 2017). For the valence dimension, the points of the scale were labeled as *Very negative*, *Moderately negative*, *Somewhat negative*, *Neutral*, *Somewhat positive*, *Moderately positive*, and *Very positive*. For the arousal dimension, the points of the scale were labeled as *Very low, Moderately low, Somewhat low, Neither low nor high, Somewhat high, Moderately high*, and *Very high*. Images were presented in individually randomized order.

After assessing the pictures, respondents were asked to answer a standard socio-demographic questionnaire, i.e., items on age, gender, education, native language, current residence region, and region where they had lived longest. After completing the survey, participants were thanked and debriefed.

- ***Results***

317 participants took part in the survey. I aimed to exclude responses of individuals whose standard deviations (SD's) across all the rated pictures were less than 0.6 (which would mean that a participant assessed almost every picture with the same score, and it would imply cheating or other suspicious attitudes). However, there were no such cases nor for valence, neither for arousal. So, no participants were excluded from the study. The mean valence and arousal ratings and the corresponding standard deviations were

calculated for each image. The distribution of the imagewise standard deviations of the means is shown in Figure 8. Valence ratings ranged from 2.09 to 5.84, showing not perfect but relatively good usage of the entire range of the scale (1-7). The mean valence rating was 4.37, somewhat above the theoretical midpoint of the scale, and the median valence standard deviation was 1.3. This correlates with the results from the original OASIS study. To check the normality of the distribution of valence ratings, I applied a Kolmogorov–Smirnov test (I compared valence results with a normal distribution where mean and SD were equal to valence ratings), $D = 0.1$, $p < .64$, i.e., the normality was confirmed. Arousal ratings ranged from 3.52 to 5.42, which showed more restriction for arousal than valence on the scale (this would have to be adapted in the programming of the app, see Section 3.2.2). The mean arousal rating was 4.29, again slightly above the theoretical midpoint of the scale. The median arousal standard deviation was 1.49 (thus higher than for valence ratings). Visually, the distribution of arousal ratings looked relatively normal, a Kolmogorov–Smirnov test for normality (I compared arousal results with a normal distribution where mean and SD were equal to arousal ratings) confirmed this impression, $D = 0.1$, $p < .7$.

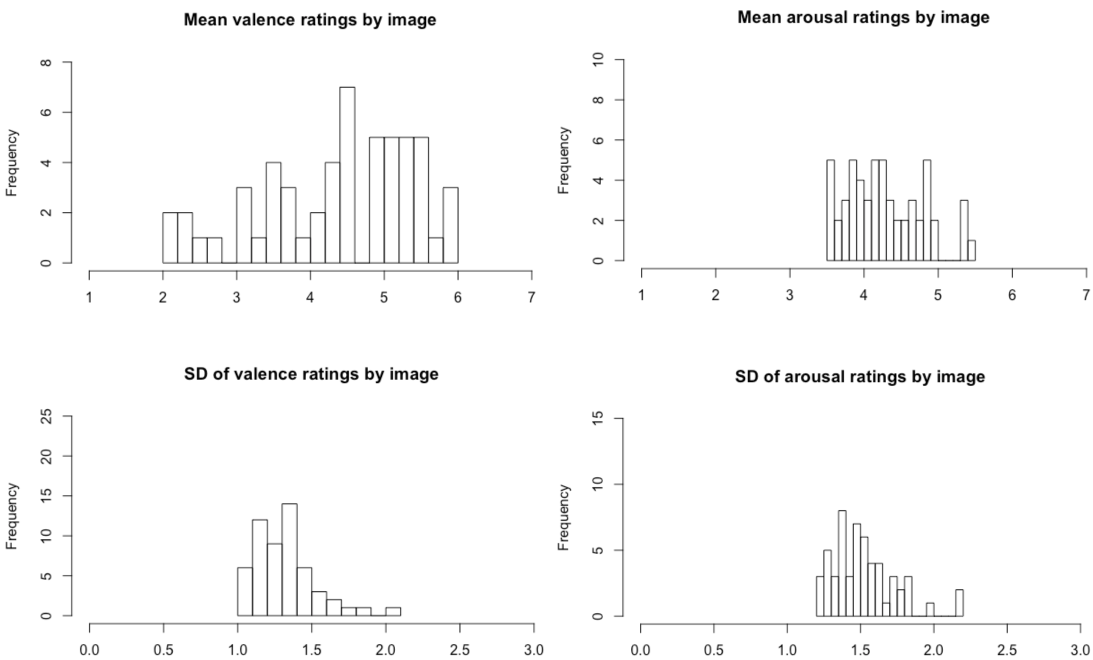

Figure 8: Univariate distributions of imagewise mean valence and arousal ratings (top row) and distributions of imagewise valence and arousal standard deviations (bottom row).

I depicted a Q-Q plot that draws the correlation between a given sample and the normal distribution in Figure 9. Both plots show that most of the points fall approximately along the reference line (the plots are just lightly tailed) - we can assume normality.

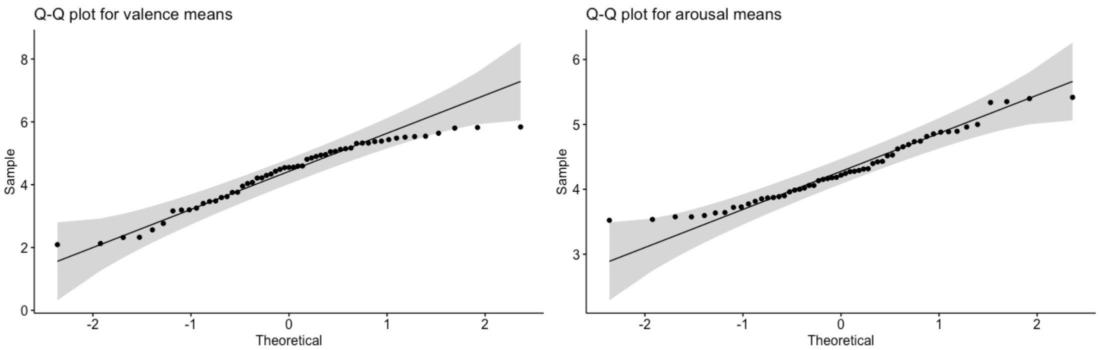

Figure 9: Q-Q plots of imagewise mean valence and arousal ratings with a 45-degree reference line. They show light tailed normal distribution of mean image ratings.

I have also compared the results for the range of arousal and valence ratings with both the OASIS paper and IAPS picture dataset: a smaller arousal range than the valence range is typical for the previous studies as well. The same worked for the higher median arousal standard deviation than the median valence standard deviation. This signifies consistency of the studies and soundness of this particular survey.

Additionally, I manually evaluated the validity of these valence and arousal ratings by looking at images that received the most highly positive and negative valence ratings and the highest and lowest arousal ratings, and at images that had the highest and lowest valence and arousal standard deviations, indicating low and high levels of agreement, respectively. The most positive valence rating (M = 5.84, SD = 1.22) was obtained for the Starfall night image, and the most negative valence rating (M = 2.09, SD = 1.37) was obtained for the image, which depicts a dark fearsome night. The picture with bright blue sky had the lowest valence standard deviation (1.04, M = 5.48), and the picture of a green sky that can be interpreted both as an ominous mystical condition or Northern lights, had the highest valence standard deviation (2.04, M = 4.54).

The highest arousal rating (M = 5.42, SD = 1.75) was obtained for the image, which depicts a thunderstorm, and the lowest arousal rating (M = 3.53, SD = 1.34) got the image, which depicts a gloomy grey sky. Picture of a calm day with a grey filter had the

Figure 10: Images with the highest and the lowest valence rating (top row), and the highest and the lowest arousal rating (bottom row).

Figure 11: Images with the lowest and the highest SD's for valence rating (top row), and the lowest and the highest SD's for arousal rating (bottom row).

lowest arousal standard deviation (1.24, M = 3.72), and the picture of a dark fearsome night (the same as for the most negative valence rating), had the highest arousal standard deviation (2.18, M = 3.64). These values support the soundness of the valence and arousal measures. The images are shown in Figure 10 and Figure 11.

The relationship between the valence and arousal ratings is shown in Figure 12. Crosses on the points represent confidence intervals (CI's) for each image. It is seen that the valence and arousal ratings create a U-form on the scale, such that arousal ratings are highest at the most positive and most negative levels of valence. This result replicates the findings in OASIS and IAPS papers. However, in my research, the form is not that expressive due to the small number of rated images.

Even though all four quadrants of the circumplex space contain a fair number of images, arousal ratings are relatively restricted under the theoretical midpoint. It may be due to strong "beauty" filters that were applied to the photos, so not that many pictures were perceived as "low energized". I considered this aspect when developing the app (see section Developing the App for more details) and designing the experiment, including the topics for discussion.

As in the previous studies (OASIS, IAPS), there was found no significant linear relationship between the valence and arousal ratings: Pearson's $r = .1$ [−.17; .36], $t(53) = .73$, $p = .46$. It may signify the independence of valence and arousal assessment. Thereby, asking each participant to rate only valence or arousal was the right strategy.

Altogether, the resulting emotional map covers 19,5% of the initial system of coordinates (measured by the edges of ratings in each direction). In comparison, the OASIS picture set covers 60% of the valence-arousal scale. It is entirely reasonable since OASIS or any other affective picture dataset uses a wide range of different images with several topics, objects, or situations to cover the affective scale as much as possible. In this experiment, I was restricted to only one picture with changing some features of it (weather condition). This obstacle, though, provided the opportunity to create a logical emotional map that does not show too explicitly the emotions of the others but sends ambiguous emotional signals to them. This approach may be compared with showing different facial expressions instead of manifesting your emotions verbally.

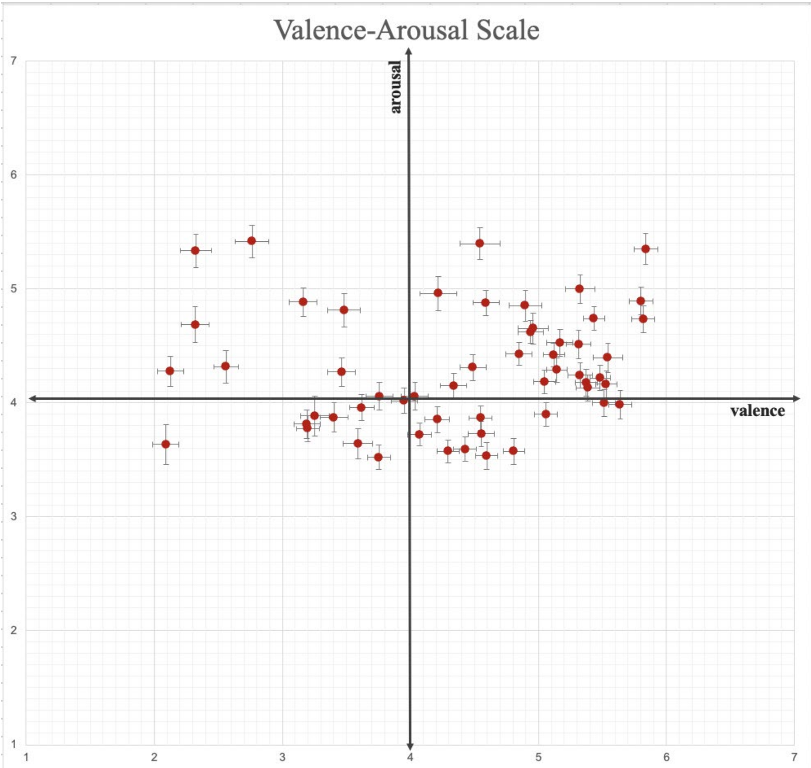

Figure 12: Emotional map: image ratings in circumplex space, with valence (measured on a 1–7 Likert scale) on the x-axis and arousal (also measured on a 1–7 Likert scale) on the y-axis.

As to conclude with this study, I would like to state that it is planned to publish the created "Emotional map" online and make it available for future research. Each picture on the emotional map was evaluated by 317 people, where valence and arousal ratings stayed independent. There are no copyright restrictions for the photos, hereby they may be used in both online and offline research studies.

3.2.2 Developing the App[6]

After the emotional map was developed, it had to be embodied into an app that participants could use for the experiment. Notably, this app is supposed to serve both purposes: to communicate the emotional state of two partners for each other (through weather images that were properly positioned in the system of coordinates) and to record this interaction. The interaction recording will be needed to measure the distance in the points of coordinates between users - before and after conversation. That is, in my

[6] The technical part of the app was developed by Ouassim Fari, MSc.

assumption, means to measure their "emotional closeness" in different time-periods of the experiment.

- ***Mapping the images to the app***

To implement the results of the research from section Emotional map: Creating Affective Picture database, the images from the system of coordinates of the study ("Emotional map") were mapped to the valence-arousal scale in the app. The images were placed in the particular points of coordinates that were corresponding to images' ratings. The area also included a confidence interval around a point. However, some photos' confidence intervals overlapped each other - it is seen in Figure 12 and is represented by the intersection of the crosses. To avoid ambiguity, I decided to keep only one from the bunch and exclude the rest. I was guided by the level of standard deviation of each picture in making the decision. Still, some images visually stood-out too obviously from their neighbors in the system of coordinates. So, I manually had to eliminate some of them for better visualisation (smooth interpolation between pictures, see below) and map comprehension. In the end, only pictures that made it through the selection were shown in the app.

- ***Filling in the gap between the points***

Still, since the pictures do not cover the whole valence-arousal scale, the gap between their coordinates had to be filled in. According to Gärdenfors' theory of conceptual spaces (see chapter Gärdenfors' Theory for Representing Information for more details), we can treat each picture as an instance of a concept, therefore if two pictures are instances of some concept, then any pictures between them should also be the instances of the same concept (concept convexity condition). One of the methods to populate the valence-arousal scale could be Voronoi tessellation. That means treating the existing images as prototypes, and any newly created image would be a replica of the nearest prototype. In this way, the whole space would be carved into areas corresponding to the original pictures (see Figure 13 with a practical implementation of Voronoi diagram to map world airport; source: www.jasondavies.com/maps/voronoi/airports).

However, this approach would supply no change of the picture while the user moves the cursor within an area of one prototype. Then, when the participant "crosses" the boundary of the previous area to another prototype - a sudden jump to a different picture occurs.

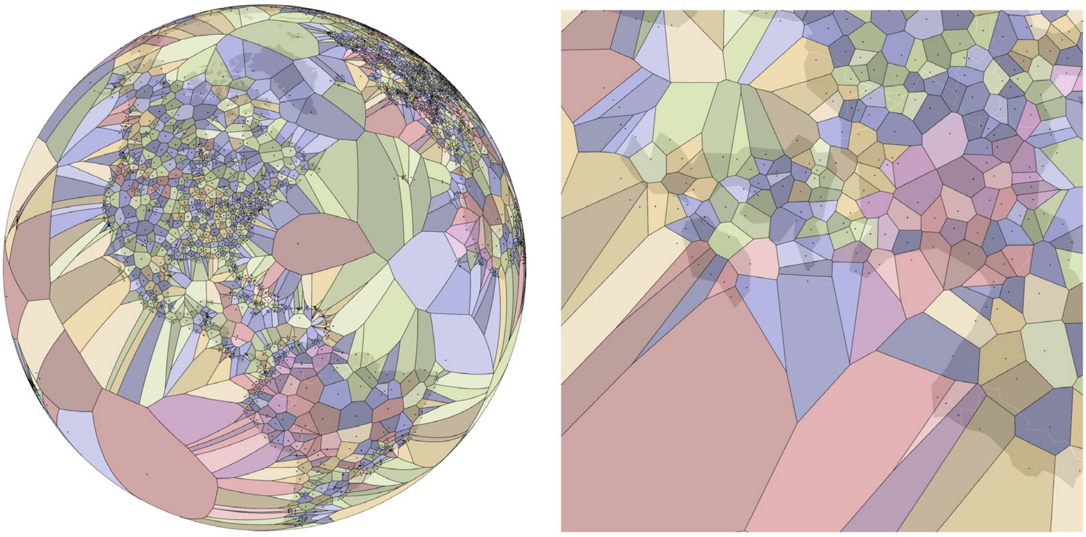

Figure 13: Voronoi tessellation for world airports map (left) and it's zoomed in version (right): each region (dyed in different colors) is closer to a particular airport than any other. Source: www.jasondavies.com/maps/voronoi/airports

In order to avoid that abrupt transition, it was decided to interpolate new pictures between existing ones. Each new image represents a weighted average of the sum of existing adjacent images. That is a weighted average of the quality dimensions around the picture in Gärdenfors terminology. To decide which images around the point to choose as a source, the area was divided into 6 radials. Only one photo per radial (the closest to the target) was picked to serve as an origin. The radial angle varied depending on the image's position on the scale, and the most viable radial size was automatically chosen each time for a satisfying outcome.

In the end, for each target image, there was a list of weighted factors (based on the source images and their distance from the target). To create a new image, we stacked all chosen pictures together with some degree of transparency (*alpha*) which depended on the weighted factor. On the top of the stack, there was always the closest adjacent picture.

In a nutshell, the value of each new picture is equal to the sum of all selected radial sources and factored by their relative distance to the target. By using this technique, the matrix with 38 original images was augmented by more than 70 000 interpolated images in-between. Additionally, in the interpolation process, different color encoding was applied to get the most natural result.

This approach provided a smooth transition from one photo to another. Therefore, the app got the effect where a thunderstorm is changing into a sunny day imperceptibly. The result is depicted in Figure 14.

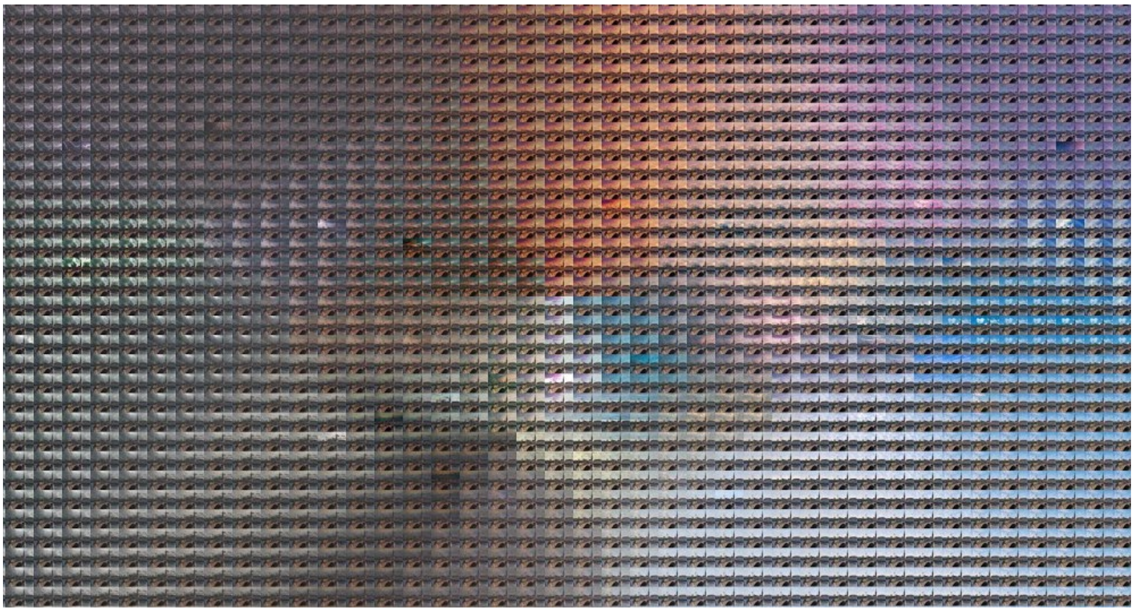

Figure 14: App matrix: original images on the emotional map augmented by newly created pictures.

- ***Cropping***

Since the resulting emotional map covers just ~20% of the theoretical valence-arousal scale the matrix was cropped by the edges of ratings in each direction. Arousal ratings had a narrower range than valence ratings, therefore the system of coordinates took the form of a rectangle instead of a square. It means I had to limit the level of possible valence and, especially, arousal expressions for the experiment discussions, and this should be taken into account.

- ***Recording***

While participants were using the app by manipulating the weather images, all updated points of coordinates were recorded on Firebase (https://firebase.google.com/). Firebase is a Google-backed application development software that enables programmers to develop iOS, Android and Web apps, and track users' activity. The recordings started when the user pressed "Start" and stopped, accordingly, when he/she pressed "Stop". The history of the coordinates was then used for the analysis of the distance between partners in each group (see Research Methodology below for details).

3.3 EXPERIMENT

3.3.1 OUTLINE OF THE EXPERIMENT

The main purpose of conducting the experiment is to see whether interactive emotional cues influence the empathic experience in computer-mediated communication. And if it is, how exactly. That means I want to investigate whether two partners who communicate in an enriched interactive environment have a better mutual emotional understanding. By emotional understanding, I mean embracing each other's emotions, the level of connectedness between partners, comfort in communication, and, in particular, emotional closeness. Emotional closeness can be realized not only metaphorically but also literally. I will explain it in a while.

Notably, the weather-emotional app that was developed for this study (see section Preparational Procedures for details) can serve not only as a communication tool but also as a measurement tool to capture the process of interaction. During the conversation in the experiment, two partners are using the app to choose the image that best reflects their emotional state. They can change the picture whenever they feel differently. Each picture has evidence-based coordinates in the valence-arousal scale. That means, if both participants reflect on and point out their emotional state by the image with coordinates, we can calculate their "emotional distance" (i.e., Euclidean distance[7]) in the existing valence-arousal system of coordinates (Euclidean space). Thereby, we can literally calculate their "emotional closeness".

The app has the possibility to record any changes of the images during the conversation with their timestamps. Hence, theoretically, it is possible to see how the "emotional distance" between partners changes over time. In this study, I explain empathy by the tendency to perceive what one's partner feels, i.e., "to empathize" means "to approach" closer to the partner's emotional state (for more details see section Empathy). Consequently, I expect that distance reduction between weather images' coordinates during the conversation signifies enhancing emotional closeness between partners, i.e., increasing empathic experience.

[7] Euclidean distance is calculated by the formula $dist((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}$ where (x,y) and (a,b) are coordinates of the points.

Since during the conversation, participants could not always react simultaneously, and sometimes could forget to change the picture in the momentum, in my analysis I used only information at the beginning and at the end of the conversation (participants got explicit reminders to adjust the picture at these time periods). I compared this information (differences in the distance) between the test and the control groups (I shortly described two groups in section Short Overview of the Method and in more details in section Materials & Methods).

"After-conversational" change in emotional closeness (differences in the distance of coordinates) was one of the parameters to compare the emotional experience in two groups. However, this way of emotional interaction evaluation is new in the field, hence, it may be questioned. Some could argue that these numbers could signify nothing but coordinates in the scale, while real participants' emotions stayed uncovered. To avoid potential ambiguity of experiment interpretation, I decided to collect additional information to support my idea. Hence, I have also recorded participants' conversations during the experiment, and joint reflection at the end of the study (before debriefing). Additionally, before and after the conversation participants were asked about their current emotional state: they had to give a brief answer in an open form in 1-2 short sentences. Moreover, before joint reflection, they were asked to fill in a short questionnaire regarding their felt level of comfort, connectedness, and understanding each other's emotions during and after the conversation. The questionnaire consisted of 4 questions on a 7-point Likert scale and 4 open questions where respondents were asked to explain the chosen rating.

To make sure that the two partners were in different emotional states before the beginning of the conversation, I gave them different inputs. Each participant was asked to watch a short video clip, where participant A watched an exciting inspiring video, and participant B watched a boring/annoying video. I described more details about the video input in section Materials & Methods.

Altogether, an experiment session took ~50 min. The structure of each session for both (control and test) groups looked like this:

1. Welcoming participants, short outline and the purpose of the experiment (5 min)
2. Introducing the app, getting familiar with it, registration for the session (5 min)

3. Watching short input video (3 min) + further instruction (5 min)
4. Answering the question about the current emotional state (1-2 min)
5. Conversation part with using the app (13 min)
6. Answering the question about the current emotional state (1-2 min)
7. Filling in the questionnaire regarding comfort, connectedness, understanding each other's emotions (4 Likert-scale questions, 4 open questions for explanation) (6 min)
8. Feedback and joint reflection; debriefing (10 min)

3.3.2 MATERIALS & METHODS

- ***Participants***

Altogether, 56 people took part in the experiment. Participants were randomly paired up and formed 28 sessions. Like in the study where I aimed to create the Emotional map, there were some shifts in the data. These shifts correlated with the first study: there were more females among the participants (74.1%), and most of the participants were from Europe (94.8% chose Europe as a current residence region and 87.9% stated that they have lived the longest period in that part of the world). I also asked participants about their native language. For obvious reasons (fewer respondents), there were fewer languages represented in the study, but the distribution was more heterogeneous: around a quarter (25.9%) of all participants spoke Russian as mother-tongue, 22.4% were Slovenian native speakers (some students of MEi:CogSci Ljubljana were asked to participate in the study), 12.1% had German as the first language, English and Slovak shared 8.6%, and 5.2% were bilingual or multilingual. Younger generations mostly took part in the main study: in the age between 18 and 34 (27.6% were in the 18-24 age group, and 60.3% were 25-34 years old). The distribution of all socio-demographic data is provided in Figure 15.

To avoid any hidden affections or dislikes all pairs that were formed for the study, consisted of random strangers. Before confirming the match, participants were asked whether the potential session partner is unknown to them. This strategy had some advantages and disadvantages. Among the advantages, besides avoiding possible preexisting emotions to the partner, I can name a high level of participants' curiosity to meet new people - this helped me a lot to recruit participants. Another positive effect, according to some participants' feedback, was greater willingness to open feelings in front of a stranger rather than a friend or colleague[8]. However, as a disadvantage, I would name high dependence on chance whether a new acquaintance will sympathize with each other, and their individual differences in the ability to connect quickly. I will mention this disadvantage in the Discussion chapter.

The participants were recruited through Facebook groups (related to expats and women communities, psychological and cognitive science groups of interests), as well as through word of mouth (acquaintances of the researcher).

- ***Design***

For the experiment design, I deliberated whether to use the between-group or within-group method.

In a within-group design, same participants take part in both (or more) conditions of the experiment. In the case of this research, it means each pair would have to participate in the experiment session twice: have a conversation while using the app with and without sending each other interactive emotional cues. According to Lazar et al. (2010, pp. 46–52), there are several advantages of this design method: I would need a smaller sample size (since the same people participate in different conditions), statistical tests that are used in the within-group design are believed to be more powerful (e.g., repeated measures ANOVA), and, since we compare different conditions on the same people, there is effective isolation of individual differences (in this research, it may be: a different level of emotional intelligence, probability of “clicking” with another person, language level, specific strategies of opening oneself to strangers, etc.). As for disadvantages, I can mention the large impact of fatigue or boredom (because of repeating the experiment) and especially the hardly controlled learning effect. Some researchers suggest randomizing the order of the experimental conditions to control the learning effect. Still, the training effect may be trickier to administrate it is rather difficult to distinguish whether participants had a richer empathic experience during their second try due to

[8] In Russian culture it is called "the effect of wagon companion" (or "chronotope of wagon companion", (Andreeva, 1988)). By this effect, people explain talkativeness and openness of travellers with their random companions. It is believed that total strangers who most likely will never see each other again will be much more frank and sincere (especially, if they are situated in an enclosed place and with a limited amount of time).

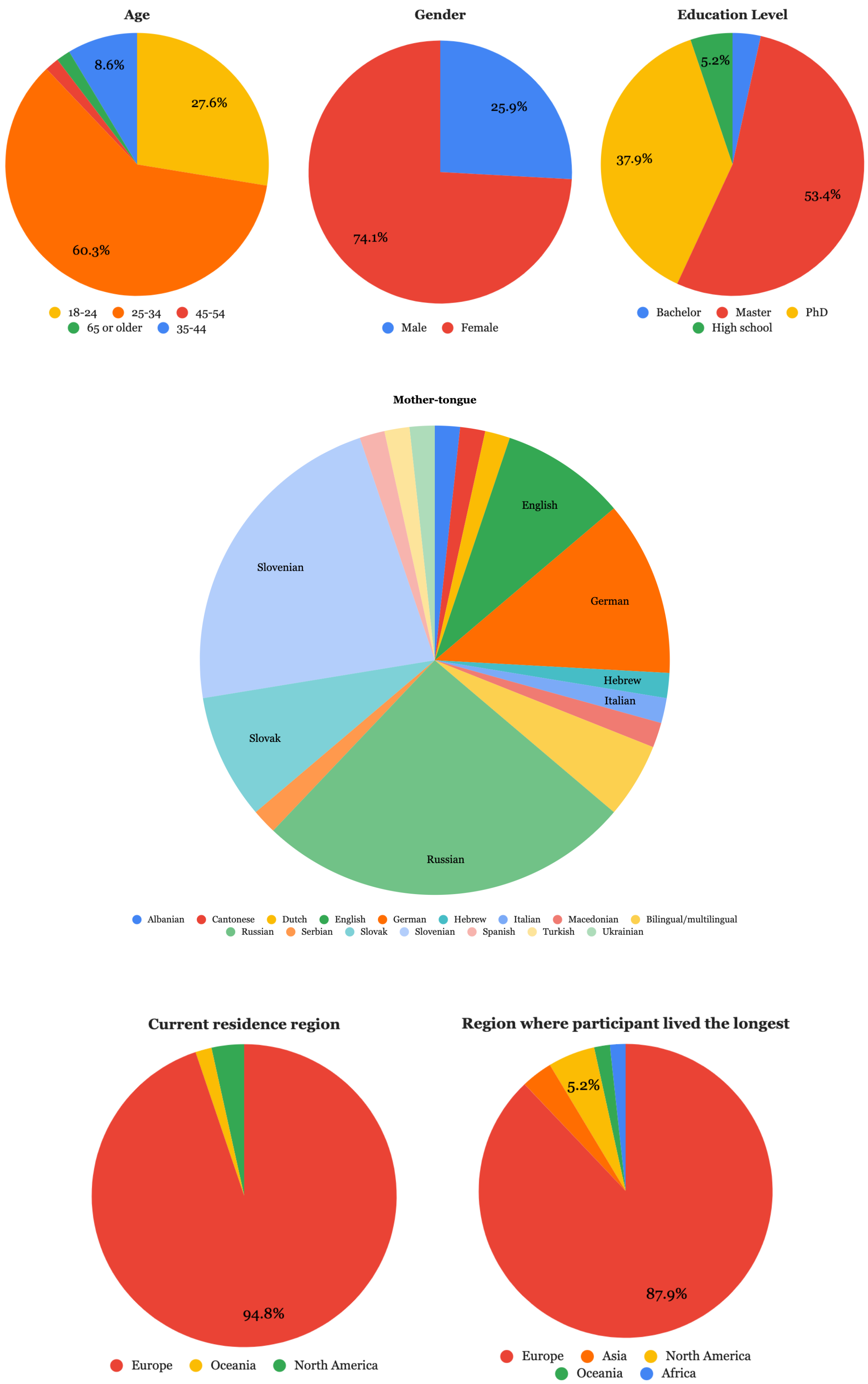

Figure 15: Socio-demographic data of respondents of the main study.

changed independent variable or because they understood the task better at the second time. And, vice versa, could the empathic experience have been poorer since participants got tired/bored after the first time, and different results had nothing to do with the measured condition? Moreover, we are dealing with social interaction, so partners could be just in a different mood if I distribute sessions in time or "catch the flow" if I organize two sessions in a row. Furthermore, in HCI practice it is recommended (Lazar et al., 2010, p. 51) to avoid within-group design while testing tools that are new to participants (like an app with a specific managing system that involves high cognitive efforts), in comparison to testing well-known features (e.g., the influence of different fonts on reading speed). This means, even though the same individuals are tested in within-group design, there are still too many specific variables to control (including individual mood on different days and the possibility to get together from the beginning), while additionally a hard problem of learning effect occurs.

In a between-group design, more participants are recruited for the experiment: they are divided into groups by the number of conditions that are tested. Which means each of them takes part in the experiment only once. In my research, I used a between-group design. I have recruited participants and divided them into two groups: test and control (see the detailed description of groups below). According to Lazar et al. (2010, pp. 46–52), this design method is cleaner than within-group design, it avoids learning effect, and facilitates better control of confounding factors, such as fatigue or habituation. There are some disadvantages as well: the researcher needs a larger sample size for the experiment (required amount of participants should be multiplied by the number of groups, i.e., conditions), it is supposed that it is harder to get statistically significant results[9], and the large impact of individual differences should be taken into account.

In my research individual differences played a big role (disadvantage of between-group design), however, I considered the problem of training effect as more important (disadvantage of within-group design). Moreover, since I used strangers as partners in sessions, repeating the experiment with the second condition in within-group design would have led to another problem: they would not be a pair of strangers anymore - they had already had a session together (so, they could be biased for the second session). As

[9] I used t-test for my analysis, but it is also common to use one-way or factorial ANOVA if the study has more conditions.

to develop this idea, follow-up mixing up the pair for the second session would anyway have brought the problem of individual differences on the scene (they would form a new pair and would communicate differently). Hereby, the research got the between-group design.

- ***Test & Control groups***

The two groups have different means of communication between partners. They are distinguished based on the presence of a particular condition: auxiliary interactive cues during their conversation. Participants from both groups use the app, reflect on their emotions during the conversation, and constantly adjust the picture that best represents their emotional states.

However, participants from the test group see the updates from their partners: on the second screen (rectangle above the agent's image) the picture is changing based on the input sent from the partner. The agent can react to it, process this extra signal, or try to ignore it if she decides so. Still, she will be influenced by this auxiliary emotional cue that is not expressed verbally. Our hypothesis is that this extra interactive channel may help in interpreting the agent's own and her partner's emotions more deeply.

In the control group, the second screen (rectangle above the agent's image) is "frozen" on a particular (neutral) image. This means, the agent sees only the changes of her own image reflected in the app (rectangle below the frozen image and above the valence-arousal scale), and the “partner's" image stays the same. Hereby, both participants have a normal audio conversation with each other while marking their emotional states in the app (these data are sent to the server for the future analysis of the results).

In both groups, participants were instructed that the main purpose of the app is to help them better reflect on their emotions. This helped to avoid possible confusion in the control group regarding the meaning of the task. In the test group, participants also understood that their images were seen by their partners. However, they were told that they were free to react or ignore the input. "*You should not feel obliged to react specifically and behave according to somebody's (your partner or researcher) expectations. What is important are your emotions at any particular moment during the experiment. Whether you feel that this information (the "emotional" images) is valuable,* *you may use it or not in your conversation*", was said to every participant in the test group.

14 sessions were conducted for the test group and 14 sessions for the control group.

- ***Platform & Tools***

Experiment sessions took place during the COVID-19 pandemic, so all processes were organized online. Interested people (potential participants) registered in the experiment, filled in the paper of informed consent, and read the instructions and overview of the experiment in Google Forms in advance. The time slots for the participants' sessions were chosen through Doodle (https://doodle.com/) - a free online meeting scheduling tool. Each participant received their ID in advance and they used it for choosing time slots. I matched people who chose the same slots and were recruited from different resources (to prevent matching friends / colleagues / acquaintances) with each other.

During the session, the participants and the researcher connected by Zoom (https://zoom.us/) - a platform for video and audio conferences. Until the joint reflection and debriefing part, all attendees communicated with turned-off cameras. This helped to avoid biases based on participant's appearance and visual signs of emotions. Video also did not distract from the app during the conversation.

The session was moderated by the researcher: the purpose of the study and the future steps were explained at the beginning. Since all attendees were situated in different places (and sometimes cities and even countries), accurate guidance was required. For this purpose, I again used Google Forms. It served both as guidance through the steps of the experiment[10] and as a recorder of participants' answers. All the links (for the web page of the app, and for the Youtube video) were also inserted in this Google Form. Overview of the "Empathic Interaction Research Study" Google Form is provided in Figure 16 and Figure 17.

[10] For the steps of the experiment, see detailed Step-by-step instructions to participants below.

To record the conversation as well as feedback and joint reflection I used the software Otter.AI (https://otter.ai/). Otter.AI is a voice conversations recorder with speech-to-text transcription based on artificial intelligence and machine learning.

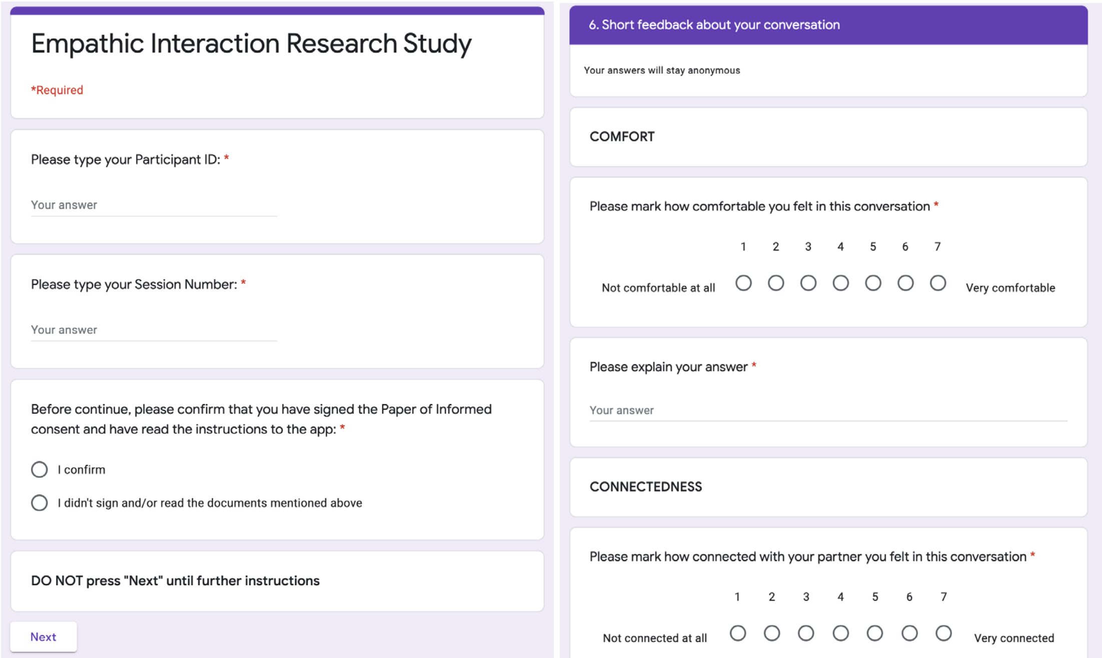

Figure 16: Data collection part of "Empathic Interaction Research Study" Google Form.

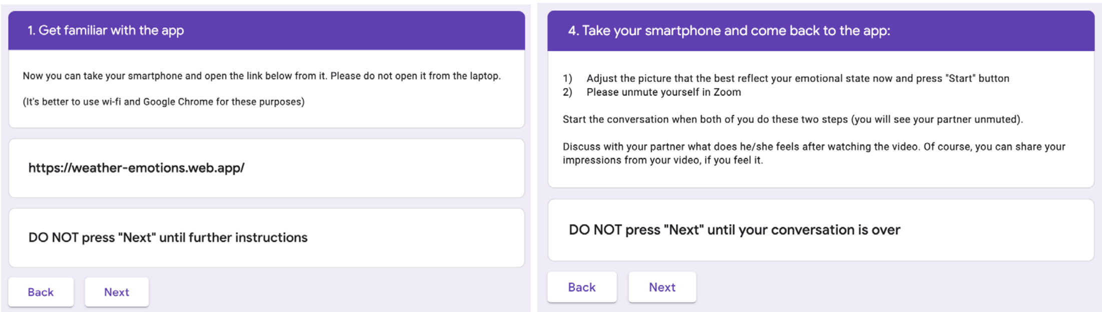

Figure 17: Guidance part of "Empathic Interaction Research Study" Google Form.

- ***Video-Input***

Partners in each pair were additionally distinguished by the video input that they have got right before the conversation. Both videos were three minutes long. The video input was prepared exclusively to try to induce a different emotional state at each of the partners at the beginning of the conversation. However, that also helped participants to start with the topic about their emotions by describing what they had seen. There was no special intention to get the participants in a particular mood, just to define a direction or the quadrant of their emotional state. That was needed to have sufficient distance between partners, so there would be some "space" for convergence.

Partner A in each pair has watched a motivational/exiting video from TV2 Danmark commercial "All That We Share"[11]. The description of the video in TV2 Danmark Youtube-channel: "We live in a time where we quickly put people in boxes. Maybe we have more in common than what we think?". The video introduces the social experiment where different groups of people (social groups) were divided by the stereotypes that surrounded them and were asked to stay in the drawn boxes. However, during the video, the host proves that people from different social groups have much more in common than we initially think, and in the end, all people from the experiment stay in one central box that unites them. This video was supposed to inspire participants and increase their level of excitement and positive spirit.

Partner B in each pair watched a boring/frustrating video of a highway, recorded from the roadside[12]. For three minutes cars were just passing in the grey weather under the rain. This video was supposed to make the participant feel bored. However, a lot of participants stated that they felt annoyed, angry, or even scared (since they expected something to happen in the video).

Even though not all the participants reported the attitude that was expected from the video, the input played the main role of "distancing" the mood of the partners.

---

[11] Youtube. 27.01.2017. TV 2 | All That We Share; [accessed 24.02.2021]. https://youtu.be/jD8tjhVO1Tc

[12] Youtube. 25.05.2016. 8hrs Cars in the Rain Highway"Sleep Sounds" ASMR; [accessed 24.02.2021]. https://youtu.be/E26A-loKmXM

- ***Step-by-step instructions to participants***

  - *Welcoming participants, short outline and the purpose of the experiment (5 min)*

– In this part, I introduce the participants to each other, explain that this study is dedicated to empathic interpersonal interaction with the help of technologies, remind that the session lasts approximately 50 minutes and that they will use both personal computer (for Google Forms, Youtube and Zoom) and smartphone (for the app). Participants are asked to open the link in Google Forms (it was sent to them 5 minutes before the start of the session together with the link for the Zoom meeting).

  - *Introducing the app, getting familiar with it, registration for the session (5 min)*

– After marking their session number and participant ID in Google Form, participants are provided with the link for the app. Through smartphones, they register for the session in the app by typing their name and session number (the same for both partners). I explain what is the valence-arousal scale and how it influences the agent's picture. I stress out that this scale is made more for their orientation and quicker finding the desired image. It means, if they feel that image that best reflects their emotional state is situated in the "wrong" place, they still should be guided by their own internal feelings, and the position is secondary here. I remind that participants should not try to follow anybody's expectations while choosing the image (neither partner's, nor researcher's) - there are no "right" or "wrong" pictures, and this app serves just as an additional channel for reflecting and expressing emotions (for the control group: "this app serves just as an additional channel for reflecting on emotions"). While I give the instruction, the participants are free to familiarize with the app.

  - *Watching short input video (3 min) + further instructions (5 min)*

– After participants get familiar with the app, they are asked to come back to Google Forms, where they are provided with Youtube videos. I explain that both of them will watch two different videos and then later they can start their conversation by discussing what they have seen.

– I remind them that after my instructions in this part, I am not supposed to interact with them and will stay as the observer in the session until the reflection part in the end. I shortly mention future steps and that I am here for them in case they have questions, but all the future instructions are written in Google Forms that they should follow step-by-step[13].

– I remind participants that they will have around 12-14 minutes for their conversation. I explain that the video will be the starting point of their discussion, but if they feel that they have nothing else to add about the topic, or the flow of the discussion moves them away, this is completely normal. They are free to have a normal dialogue like two people who have just met in ordinary circumstances with one exception: I ask them to try to avoid too much small talk and stay in the emotional/inner world level of discussion. "Try to stay in a deep level of your conversation, not superficial small talk," was one of my phrases in the instruction.

- *Answering the question about the current emotional state (1-2 min)*

– After participants have watched the videos, Google Form asks them to write in one-two sentences about their current emotional state. It helps me to compare how much their subjective reports correlate with the coordinates in the scale.

- *Conversation part with using the app (13 min)*

– Google Form asks the participants to take their smartphone (they have already been logged in in the session), adjust the picture according to their current emotional state, and unmute themselves in Zoom (this would help the agent to understand that her partner is ready). They start the conversation when both of them have done these two steps.

– Around 20-30 seconds before time is up, I notify the participants and remind them to adjust the picture in their app for the last time.

- *Answering the question about the current emotional state (1-2 min)*

– After participants have finished their conversation, Google Form asks them to write in one-two sentences about their current emotional state. It helps me to compare how much their subjective reports correlate with the coordinates in the scale.

[13] In each new section of the "Empathic Interaction Research Study" Google Form above the "Next" button it is stated whether the participant should go further or should she wait for a particular action / instructions (see Figure 17 above).

- *Filling in the questionnaire regarding comfort, connectedness, understanding each other's emotions (4 Likert-scale questions, 4 open questions for explanation) (6 min)*

– There is a short questionnaire with four Likert-scale questions:

- Please mark how comfortable you felt in this conversation;
- Please mark how connected with your partner you felt in this conversation;
- Did you feel that your emotions have been understood?
- Did you feel that you understood the emotions of your partner?

– Participants are also asked to explain their scores (answers) in an open form (see Figure 16 above as an illustration).

– It is stated that these answers will stay anonymous to their partner or anybody else.

- *Feedback and joint reflection; debriefing (10 min)*

– After the participant finishes with the questionnaire, she should unmute herself in Zoom and wait for her partner. When both of them are ready, I offer to turn on the cameras at this point, all 56 participants agreed to do so.

– Participants are asked to share any feedback regarding the experiment - from both organizational point of view, and their feelings about the experience. They can also exchange their feelings or thoughts with their partner after seeing him/her.

– After the joined reflection, I do a short debriefing by describing what I measure in this experiment, what is the difference between the two groups, and what is my hypothesis. I thank the participants and offer them to share the results if they are interested.

3.3.3 RESULTS

In this experiment, I was interested in three parameters to answer the research question: the dynamics of participants' points of coordinates before and after conversation, their subjective report based on the score of Likert-scale, and qualitative observation thanks to participants' answers and joint reflections. I compared the obtained data in all three categories for the test and the control groups and made some conjectures based on the results.

- ***Statistical analysis of the results in the system of coordinates***

– Approachment comparison

In this part, I compared the change of distance (before and after conversation) in the valence-arousal scale between the test and the control group (i.e., how much do partners approach each other emotionally at the end of the talk). That means I subtracted the distance between partners at the end of the conversation from the distance at the beginning. The resulting number I have called the approachment. All the coordinates had a decimal format for this analysis (i.e., from 20.9 to 58.4 - restricted by the minimum and maximum values of the pictures' coordinates). The mean numbers per group are represented in Table 1.

Group	Mean approachment	Distance at the beginning	Distance at the end
Control	3.16	10.86	7.70
Test	10.14	16.98	6.84

Table 1: Mean results in the distance of coordinates for test and control groups. Coordinates for the analysis were represented in decimal numbers.

Welch's two sample t-test showed a significant difference between the test ($M = 10.14$, $SD = 7.08$) and the control ($M = 3.16$, $SD = 8.49$) groups in approachment, $t(23.47) = 2.3$, $p = .03$.

However, I noticed that such a big difference is caused by a higher distance at the beginning of the conversation for the test group ($M = 16.98$, $SD = 6.75$) compared to the control group ($M = 10.86$, $SD = 6.99$). In general, partners B from the test group were more negative and indignant (higher level of arousal and a lower score for valence) at the beginning which influenced the starting distance.

Even though at the end of the conversation the distance for the test group ($M = 6.84$, $SD = 4.14$) is smaller than for the control group ($M = 7.7$, $SD = 3.36$), t-test showed no significant result, $t(24.59) = -0.59$, $p = .56$. This means we cannot make a conclusion that different condition (the presence of auxiliary interactive cues) helps the partners approach each other significantly closer.

Still, I would assume that a bigger sample size could help in answering this question (especially, since the mean distance for the test group at the end was smaller than for the control group) - 4 pairs per group is quite a small sample size for such an elaborate human task with a big influence of individual differences.

– Participants' mood path comparison

However, there is an alternative view on the data: I depicted the "paths" of participants in the test (Figure 18) and control (Figure 19) groups. The arrows show the shift of participants' mood from the beginning to the end of the conversation (no intermediate steps). Each color signifies the path of the partners in one pair. Since the paths are highly multidirectional, there is a possibility that pure numbers will not tell us the insights regarding different conditions.

Yet these illustrations show some interesting behavior that can be divided into several categories. Even though some partners tended to approach each other (some successfully, some missed the "target" - see Figure 20 for control and test groups), there were additionally other strategies, e.g., "catch-up" game (see Figure 21). In a "catch-up" game an agent approaches the initial emotional state of his/her partner, but the partner in the meantime "goes away" in the other direction (but not in the direction of the agent). The sample size in this study is rather small to make a serious scientific conclusion, still, I have noticed only one pair who did not tend to approach each other at all in the test group (however, they both were moving in parallel to the bottom-left - negative-calm - quadrant). In the control group, I have found at least ⅓ of pairs who moved in different directions and none of the partners tried to approach (see Figure 22).

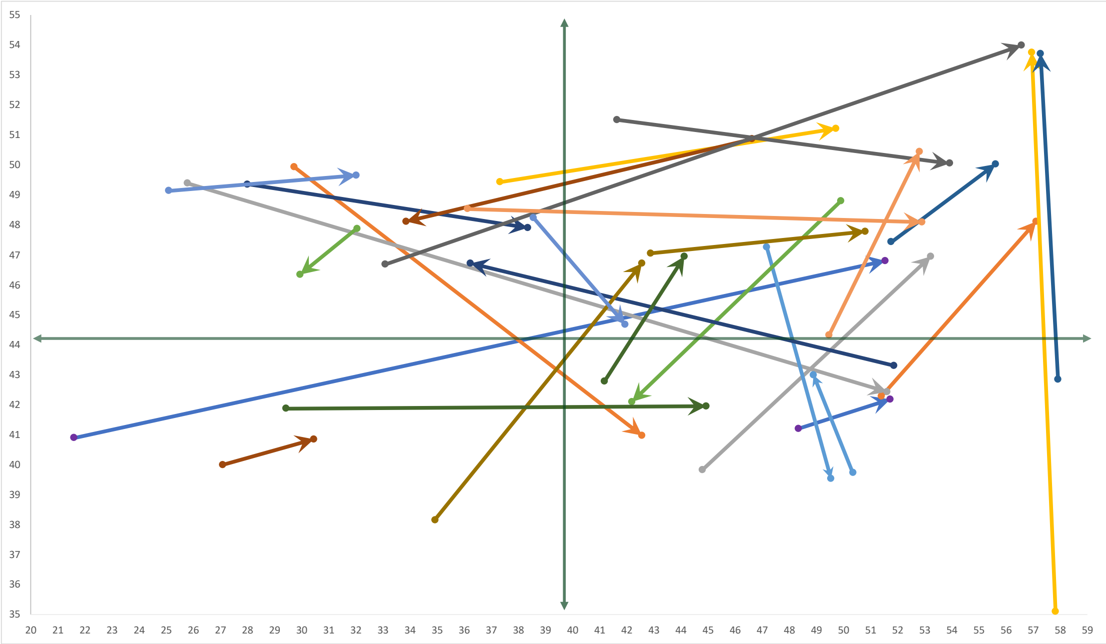

Figure 18: Partners' mood paths at the beginning and the end of conversations in the test group. Arrows show the direction of the participants' emotional states. Each color signifies the path of the partners in one pair.

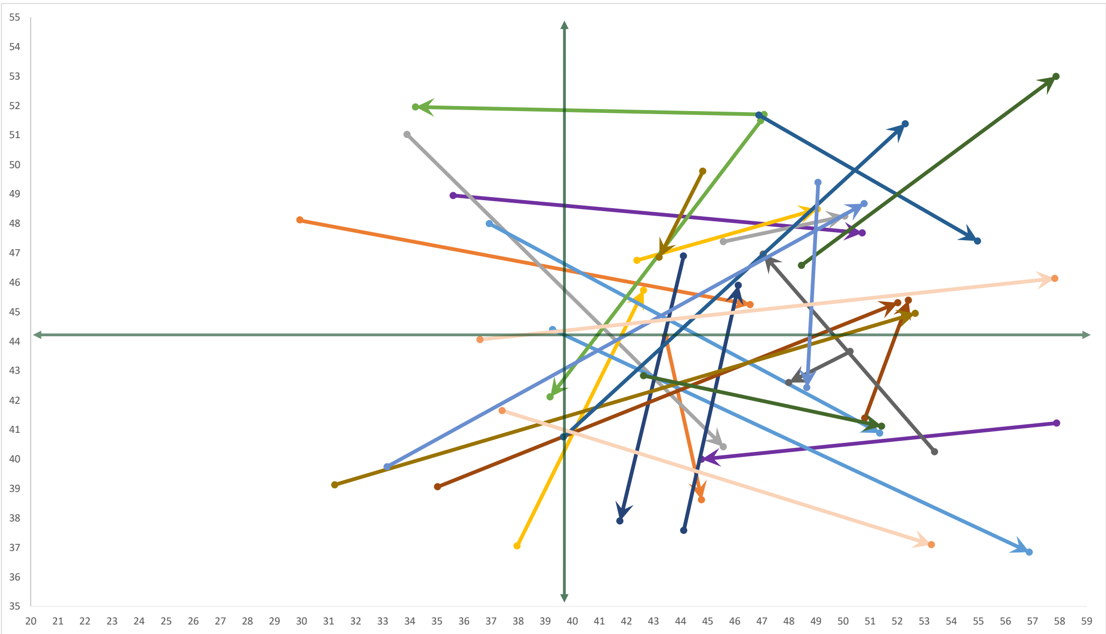

Figure 19: Partners' mood paths at the beginning and the end of conversations in the control group. Arrows show the direction of the participants' emotional states. Each color signifies the path of the partners in one pair.

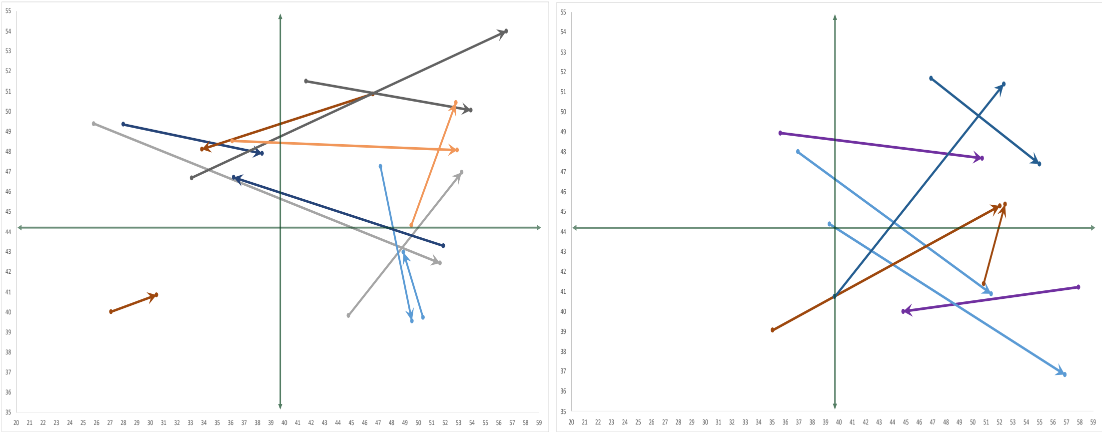

Figure 20: Approaching strategy paths for the test group (left) and the control group (right). The system of coordinates was compressed by width for better visibility.

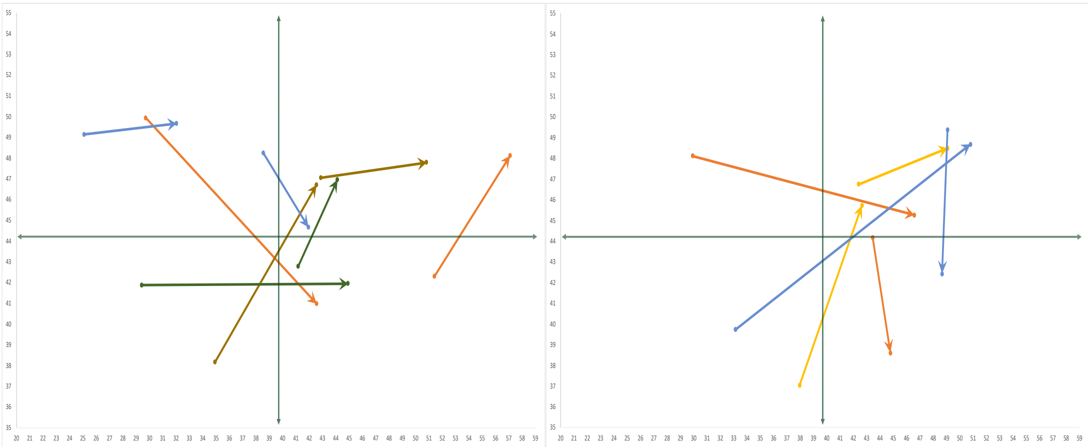

Figure 21: "Catch-up" game strategy paths for the test group (left) and the control group (right). The agent approaches the initial emotional state of the partner, while the partner "goes away" in the other direction (not in the direction of the agent). The system of coordinates was compressed by width for better visibility.

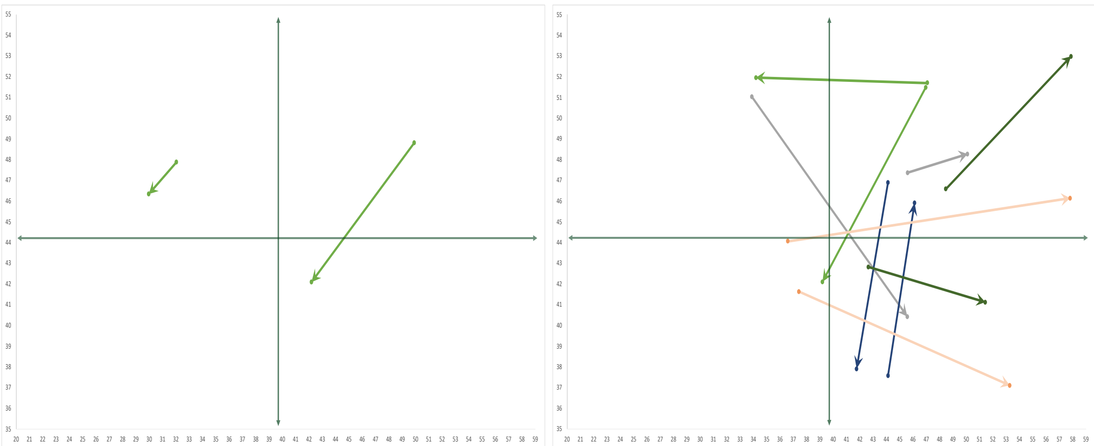

Figure 22: Different directions paths in the test group (left) and the control group (right). The system of coordinates was compressed by width for better visibility.

Another interesting behavior for the test group pairs was the synchronous approach to the top-right corner (positive-excitement quadrant) that was not noticed for the control group. This behavior of the test group participants is depicted in Figure 23. The behavior can at some level illustrate interesting findings in the levels of valence and arousal compared separately in both groups.

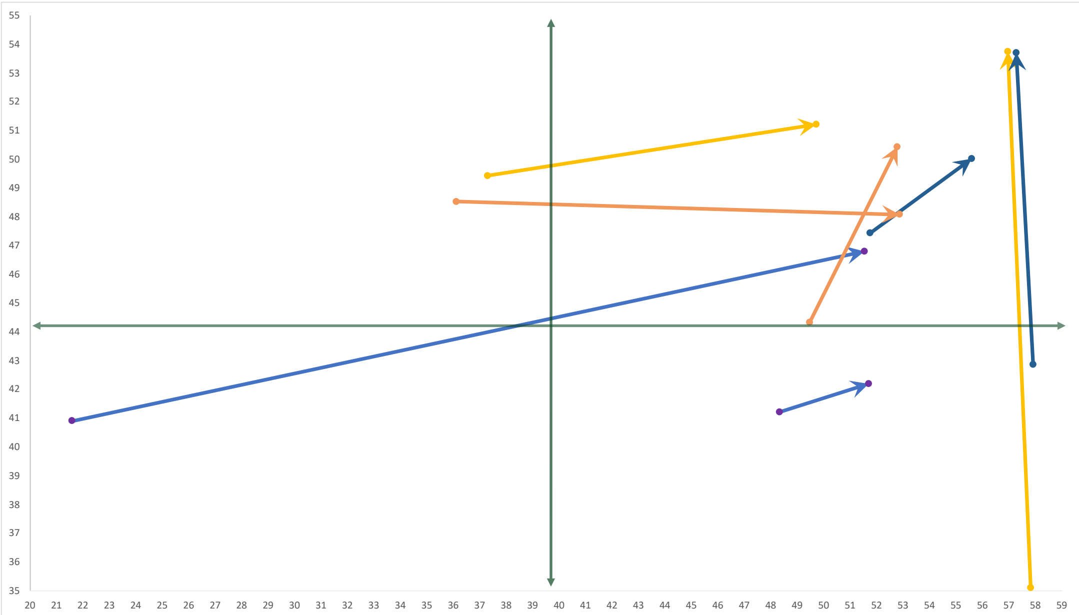

Figure 23: Partners' mood paths (in 4 out of 14 pairs) that tend to synchronously approach the top-right corner (positive-excitement quadrant) in the test group.

– <u>Arousal and valence level comparison</u>

Even though there is no significant difference between groups in changing their arousal level, it is seen that its mean level increases for the test group (M = 45.13, SD = 4.4 in the beginning, and M = 46.81, SD = 4.05, at the end). The numbers are represented in Table 2.

Group	Difference	Mean arousal level at the beginning	Mean arousal level at the end
Control	-0.18	44.80	44.62
Test	1.68	45.13	46.81

Table 2: Mean results in changing the arousal level for test and control groups.

This happened due to the changed arousal level of partners A (those who watched the positive-excited video). Partners B had a different mood path: they were highly excited already in the beginning because of the "annoying and stressed" video they have watched, so after the conversation, their high arousal emotional state only changed the valence - from negative to positive. Hence, for partners B, changes both in arousal and valence were not statistically significant.

However, partners A in the test group, while staying in positive valence, increased their level of excitement (M = 43.89, SD = 4.33 in the beginning, and M = 46.69, SD = 4.46, at the end), albeit partners A in the control group decreased it (M = 44.96, SD = 4.54 in the beginning, and M = 42.62, SD = 4.11, at the end). The numbers are presented in Table 3. The difference in the arousal reaction for partners A between groups is significant, $t(24.98) = 2.15$, $p = .04$.

Group	Partner	Difference	Mean arousal level at the beginning	Mean arousal level at the end
Control	A	**-2.34**	44.96	42.62
Test	A	**2.80**	43.89	46.69
Control	B	1.98	44.64	46.61
Test	B	0.56	46.37	46.93

Table 3: Mean results in changing the arousal level for test and control groups divided by partner type.

Group	Partner	Difference	Mean arousal level at the beginning	Mean arousal level at the end
Control	A	2.20	45.73	47.94
Test	A	2.02	46.42	48.44
Control	B	10.76	38.88	49.65
Test	B	10.66	34.41	45.07

Table 4: Mean results in changing the valence level for test and control groups divided by partner type.

In the valence dimension, there are no significant differences in mood changes neither by groups nor by partner types. In all categories, the mean valence score increased at the end of the conversation. Valence scores for partners B were much lower than for partners A due to the differences in the input (which was the initial idea for the video input), so it increased dramatically to reach more or less the same level as partners A had.

This finding may have an interesting impact and worth researching further: it emphasizes that in interpersonal communication, agents with negative mood strive to reach the emotional state of a more positive partner. While agents with high valence scores do not tend to decrease it to approach closer to their gloomy partner. Further research in this direction (or replication with a bigger sample size) may help to answer interesting questions regarding the nature of empathy in general, not just in computer-mediated communications.

- ***Statistical analysis of the questionnaire responses (Likert-scale)***

After the conversation task, participants answered four questions related to comfort, connectedness, perceiving being understood, and perceiving understanding emotions in the pair. The questions were asked in Google Form and stayed confidential.

Overall feedback regarding four questions was fairly high: the mean scores were between 5.46 and 6.07 on 7-point Likert-scale among all questions and both groups.

Comparison of the questionnaire responses between the test and the control groups showed no significant differences between the test and control group (both by t-test and Wilcoxon test).

However, all results for the test group are slightly higher (besides connectedness that is equal for both groups). Again, to prove or reject the hypothesis of better empathic experience, more data (a bigger sample size) should be collected. Table 5 represents the mean responses per each question in each group with the p-values based on the T-test and Wilcoxon test.

Question	Test group	Control group	T-test	Wilcoxon test
Comfort	5.929	5.679	0.434	0.473
Connectedness	5.571	5.571	1.000	0.944
My emotions were understood	6.071	5.857	0.531	0.972
I understood partner's emotions	5.750	5.464	0.282	0.391

Table 5: Mean results for questionnaire responses for test and control group with

In the analysis of the questionnaire responses there are again differences between partners A and B: the level of reported connectedness and understanding partner's emotions were higher for Partner B (M = 5.96, SD = 0.74 for connectedness, M = 5.9, SD = 0.83 for understanding partner's emotions) compared to Partner A (M = 5.18, SD = 1.39 for connectedness, M = 5.32, SD = 1.06 for understanding partner's emotions) altogether and in groups separately (see Table 6). Welch's two-sample t-test proved the significance of these differences both for connectedness ($t(41.33) = -2.64$, $p = .01$) and for understanding partner's emotions ($t(51.19) = -2.25$, $p = .03$).

In Figure 24 it is seen that partners B in both groups have a shorter range of scores, and the responses stick to the top scores for these two parameters (connectedness and understanding partner's emotions). Surely, these distributions should be taken into account with bigger sample size, but even with 56 responses, they give the researcher a nice overview of the potential directions.

The contained result may be due to the content that participants had to explain to each other in the beginning: a more elaborate scenario of partners A helped partners B to better grasp the idea and get some assumptions about their partner's emotions. The story about the video from partners A entertains, and there are a lot of ideas and beliefs to share that are emotionally contagious. Perhaps, the emotions from the video with poor content (cars passing by on the highway) are more difficult to communicate.

Question	Test / A	Test / B	Control / A	Control / B	T-test	Wilcoxon test
Comfort	5.86	6.00	5.71	5.64	.91	.90
Connectedness	**5.14**	**5.93**	**5.21**	**5.93**	.01 *	.02 *
My emotions were understood	5.64	6.50	5.86	5.86	.21	.18
I understood partner's emotions	**5.50**	**6.00**	**5.14**	**5.79**	.03 *	.04 *

Table 6: Mean results for questionnaire responses per group and partner. T-test and Wilcoxon test are applied for comparison within partner types (A and B), not within groups.

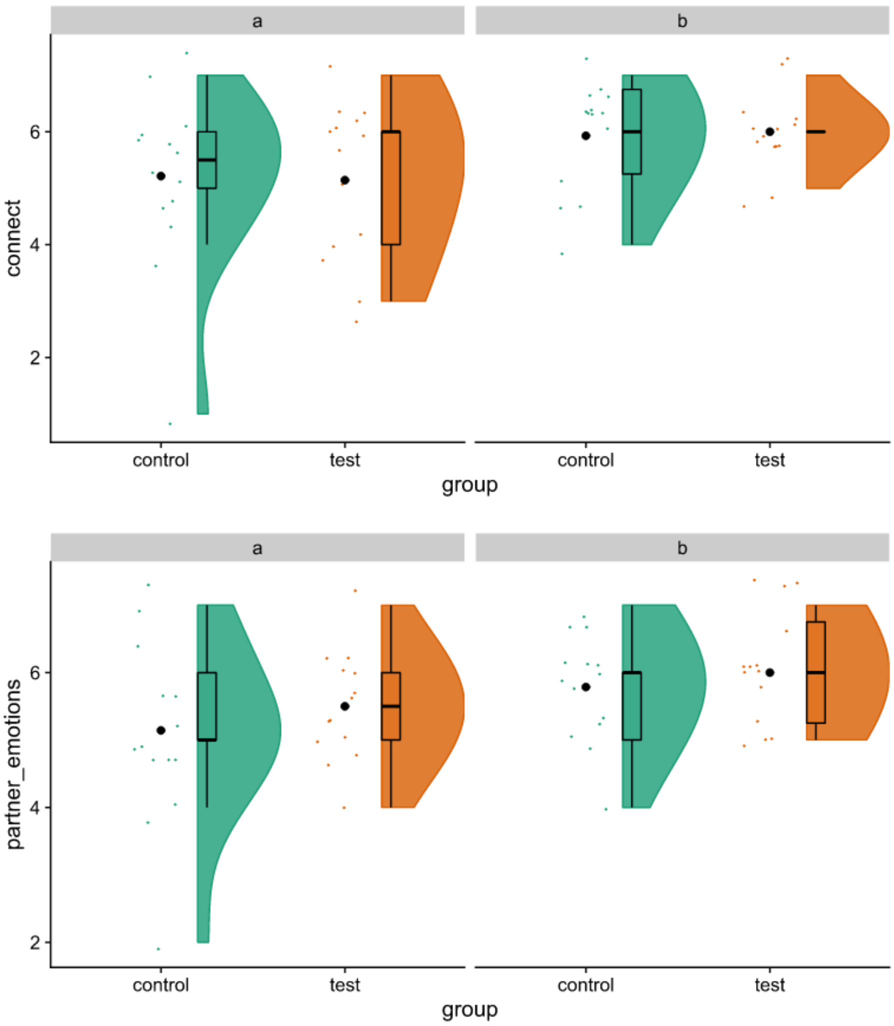

Figure 24: Violin plots showing distribution of participants' responses for the questions of connectedness (top) and understanding partner's emotions (bottom) per group and partner type. Color dots are responses, black dots are means.

- ***Qualitative observation***

There was no explicit intention in using qualitative methods. Initially, I planned to use qualitative observation to help interpret the quantitative results from the points of coordinates and questionnaire responses parts. However, the numerical findings for these data were not that impressive to make confident conclusions. Hence, the qualitative part may shed some light on our research.

There were several resources for qualitative observations: recorded scripts from participants' conversations, their reports regarding emotional state before and after conversation, responses on open-ended questions on the questionnaire, and joint reflection/feedback part. I found that the most interesting for the study were short written reports before and after conversation, questionnaire, and feedback part. No specially structured coding was used for the analysis, instead, I made very personal observations on participant's words.

Short written reports helped me to understand better the shifts in participants' mood and also to juxtapose them with the initial and final points of coordinates from the app. I have transferred the obtained result on Figure 25 (for the test group) and Figure 26 (for the control group). Dots are the points of coordinates in the valence-arousal scale, where darker ones represent the initial emotional state, and brighter ones - the final emotional state at the end of the conversation. Arrows indicate the path from one to another (no intermediate points). Purple-to-pink paths are for partners A, and orange-to-yellow paths are for partners B. Inconsistent or illogical (to the coordinates) reports were taken into red frames.

From first sight it is clear that the initial emotional states of partners B in the test group are much further from partners A, hence test group users (B) overcome a greater distance than control group participants. Interestingly, visually both partner types in the test group tend to reach a positive-excited quadrant of the scale. In the control group, the picture shows that participants reach a positive-excited quadrant as well, but also a positive-calm quadrant. Many more participants are concentrated only on the right side of the scale.

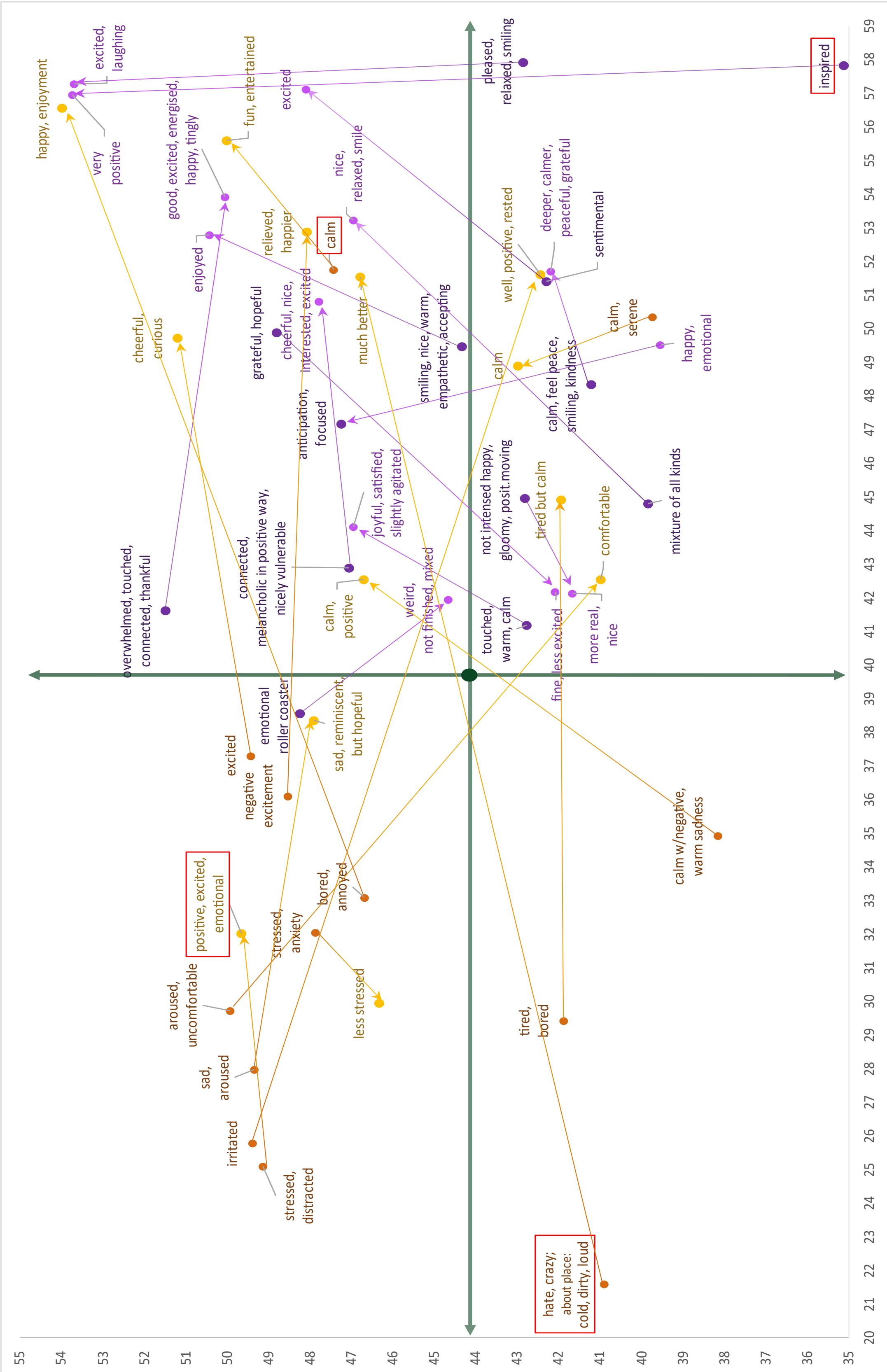

Figure 25: Test group. Juxtaposing the participants' "mood paths" – initial (darker) and final (brighter) points of coordinates connected by arrows – with the reposted subjective emotional states. Purple-to-pink paths - partners A, and orange-to-yellow paths - partners B. Inconsistent or illogical reports were taken into red frames.

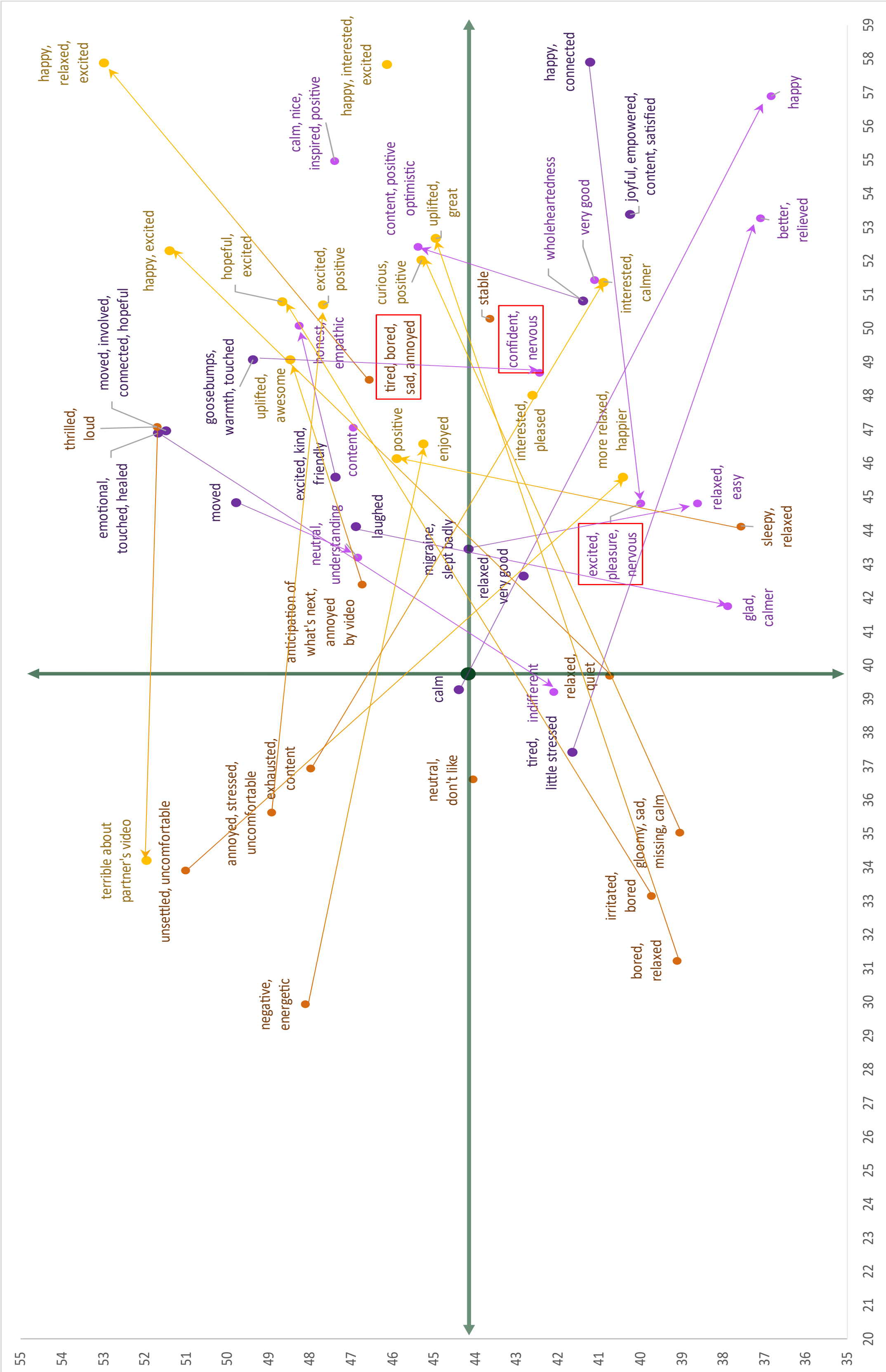

Figure 26: Control group. Juxtaposing the participants' "mood paths" – initial (darker) and final (brighter) points of coordinates connected by arrows – with the reposted subjective emotional states. Purple-to-pink paths - partners A, and orange-to-yellow paths - partners B. Inconsistent or illogical reports were taken into red frames.

In the **questionnaire responses part**, it was quite often mentioned the fact that partners did not know each other, that they talked with total strangers, and that they have never seen the face of their partner before. Sometimes it was stated as a disadvantage ("*It is very nice to feel this connection taking into consideration you do not know and cannot see the other person*", "*For talking to a stranger, we had some pretty interesting discussion and I really enjoyed it. Not being able to see the other person I'm talking to or know anything about them makes me feel a little hesitant to feel 100% comfortable.*"), sometimes as a benefit ("*It was nice to talk with an unknown person about my feelings*", "*It was an amazing experience to share emotions with a stranger. Felt like Speed-Dating*", "*Great idea, I liked talking to a random stranger about feelings - should be done more often*", "*It was very interesting to speak to a "stranger" about feelings and emotions over a voice call. I thought that not seeing the person would present more barriers for connecting over the conversation, but it really didn't, there was a special form of connectedness.*"). Nevertheless, the fact of talking to a stranger influenced participants' overall impression of the experiment considerably. This should be taken into consideration for interpreting the results.

Many participants mentioned the no-video format of conversation. An interesting aspect is that partners from the control group (no interactional cues) reported it as a negative aspect more often ("*I was a bit unsure though because I did not see her face and could not see her facial reactions*".), while partners from the test group found it as an advantage or simply an interesting experience ("*It was pleasant not to see his face, so I felt more comfortable to talk*", "*Because we did not see our faces, we had to focus more on our voices*", "*I like this experiment even more without the faces and the physical proximity!*"). This may give us a hint that test group users who had seen auxiliary interactive cues, compensated "missed" visual stimuli of the partner's face by partner's shared emotional state (represented by the weather pictures).

Moreover, test group users often mentioned the beneficial effect of the app with exchanging pictures ("Since I could see his emotions with the App it was nice to imagine how he feels right now", "And that I could see in the app how he feels made me feel he is closer to me", "I think the picture was very helpful to understand the state of mind and emotions. I couldn't however grasp it from the voice and intonation of my partner"). For some of the participants, this experience was quite vulnerable, and more sensitive compared to a conversation in real life: "It was quite interesting but I didn't expect to feel this kind of naked. It felt like the conversation itself was not that deep, but it still felt deep for me because I had to show my emotions, the whole time."

During one of the sessions, technical problems occurred and the app of one of the participants got "frozen". It was reported by her partner and influenced her overall impression: "*Verbally, I felt very connected, but I kept looking at her picture and felt like it wasn't changing at all, even though I was feeling huge changes in myself.*" This case emphasizes the important impact of the app on partners' impression of conversation.

One of the findings in the analysis of questionnaire responses was better connectedness and understanding partner's emotions among partners B. My assumption was that the video content for partners A was more elaborate and its detailed explanation helped partners B to better grasp the idea. This assumption correlates with participants' feedback among partners B. They mention in the majority of reports that the video description was very interesting and how much their partners (A) were open to discuss their impressions from the video. It helped them to understand their partners' feelings and being closer to them ("*I respect the openness of my partner to easily describe a video, so it brought a feeling of proper connectedness*", "*Even though I was initially irritated by my own video, once I heard about her video I felt like I could see what she had seen, and it improved my emotional state.*"), and, vice versa, partners A was more confused about the content of partners B video, and it influenced their impression from the conversation ("*I felt just slightly shy when I didn't know what to say about her video*", "*I am not quite sure that I completely understood the emotions of my partner, because the video she described had basically no story*."). Thereby, these short reports correlate with quantitative findings in the questionnaire and let me make a conclusion that the assumption was correct.

In general, most of the participants gave very generous assessments of the experiment, though I did not notice a big difference in reviews between test and control groups ("I *was surprised how easily he opened up, which made me want to do the same. He talked about the things he feels nowadays which are close to me.*", "*Before the start, I was exhausted and angry. Now I am feeling even better than before the experiment. Happier and more energetic.*"). One of the participants pointed out that this type of experience may influence one's perception not in the current moment, but for the future or for the past: "*During the conversation, my emotions didn't really change. More what I am talking about - my emotions (regarding) the past had changed.*"

Participants emphasized the importance of "being rightly matched", the topic that is concentrated around emotions, and an openness-stimulating design of the session. Some participants even willed to exchange contacts after the experiment and "meet" again (and they did) or repeat this experience again with new strangers: "*It was an amazing experience, meeting educated strangers like this, no video, and I would love this to be an app for real or something we could use to meet strangers".*

Even though participants generally had a positive experience, this communication still looked like a "bridge" for normal communication in real life: "*It was a lovely conversation that I'd love to do again - maybe over a cup of coffee.*". However, there was no goal to substitute real conversation: the main purpose of this study was to find out how we can enhance our communication with other people in the light of empathic experience. From this point, one of the participants' reports represents the biggest value. In the last question of feedback "Please tell what else you want to share with me regarding this conversation" he replied: "*One is forgetting about the app while having a good conversation.*"

4 DISCUSSION

4.1 MAIN FINDINGS AND INSIGHTS

Even though the obtained data from the experiment did not show a generally expected result - there were not many parameters with significant differences between the test and the control groups - some interesting insights should be taken into consideration. The "path" that test group partners made (the changed distance in coordinates) during the conversation is significantly higher than for the control group. And, despite the fact that it is mainly due to the starting positions of the partners, the final mean distance between the test group compared to the control group is still smaller. This could mean that we should not count out our hypothesis. These data may be a premise for further research in this direction - preferably with bigger sample size or a slightly different research design (see Limitations and Suggestions for Improvement section below).

Qualitative observation additionally speaks in favor of my hypothesis: participants' reports show that test group partners used auxiliary interactive cues (shared emotional state represented by the weather pictures) to compensate the “missed” visual stimuli of the partner's face. This conjecture is drawn since the test group mostly reported the positive reaction on turned-off cameras during conversations (e.g., "*It was pleasant not to see his face, so I felt more comfortable to talk*") compared to the control group where participants felt uncomfortable without seeing partners' face. Hence, we can make a conclusion that auxiliary interactive cues play a relevant role in empathic experience in computer-mediated communication. This conclusion is in accord with other parts of the reports where test group participants showed their favor to that app feature: "*And that I could see in the app how he feels made me feel he is closer to me.*" and “*It was quite interesting but I didn't expect to feel this kind of naked. It felt like the conversation itself was not that deep but it still felt deep for me because I had to show my emotions, the whole time.*”

Another interesting insight from the experiment is the so-called "mood paths" - the trajectories that participants followed from the starting to the final points of coordinates (intermediate steps were not analyzed). In section Participants' mood path comparison I assumed that the numbers do not show such extreme results due to the different behavioral approaches (trajectories) of the partners. Some partners were trying to approach each other (several successfully, others failed), but some were "playing" catch-up game (where an agent approaches the initial emotional state of his/her partner, but the partner in the meantime "goes away" in the other direction). The latter approach may signify that one partner was more engaged/responsive than the other but there was still some emotional understanding or even sympathy at least from one side. These findings may not be noticed by just looking at summarizing statistics.

I have also noticed that in the test group only one pair did not tend to approach each other at all, while in the control group there were at least ⅓ of pairs who moved in different directions and none of the partners tried to approach. The sample size is too small to do scientific conclusions from these observations, however, it may signify better emotional understanding among test group partners rather than among control group partners.

There may be a different research design or even research question to get more insights from partners' trajectories in the valence-arousal scale. Nevertheless, I propose that "mood paths" may be a valuable approach to study interpersonal communications and this method is definitely worth investigating more.

Another insight is related to an unexpected finding: the significant difference between partner types. Partners A in the experiment were participants who watched a video input with the exciting and inspiring storyline, while partners B were participants who watched how the cars were passing by on the rainy highway (boring/annoying video). Several comparison methods (points of coordinates and reported answers for questionnaire) showed that partners B had the tendency to emotionally approach closer to partners A (i.e., perceive better partner's emotions). They tend to increase their mood (increase valence level) more than partners A tend to decrease theirs (moreover, mean valence level of partners A even slightly increased). It suggests that in interpersonal communication, agents with negative mood strive to reach the emotional state of a more positive partner. While a positive partner does not tend to decrease his/her. This finding may be worth further investigation and be useful for studying empathic interactions in general, and methods in computer-mediated communications can help researchers with it.

4.2 LIMITATIONS AND SUGGESTIONS FOR IMPROVEMENT

It is always ambitious to try to embrace the broad notions in one research. In this case, it was ambitious to try to put emotions and empathy into specific numbers or schemas. The empathic interaction process may be much more complex than the mockup I have tried to investigate in my research. However, working with such complex terms requires simplification of vast concepts so they can be examined. Therefore, it is still a useful starting point for more elaborate studies.

Some of the limitations that I have already mentioned earlier (see Design section) are factors of individual differences between participants, differences in partners' matches (i.e., how well random strangers will "click" with each other), and participant's particular mood in a moment of experiment. The latter I tried to minimize by giving specific video inputs for the partners. Regarding the individual differences and differences in participants' matches - if somebody decides to replicate the study - the solution would be to simply choose a large enough sample size so that these effects would be averaged out in both groups. An alternative would be to recruit partners who are already good friends and, therefore, apply the within-group design. This can help with minimizing (but not completely avoiding) differences in partners' "clicks" and solve the problem with individual differences. Still, the researcher would have to take into account that this study would be dedicated to empathic interaction through technologies among good friends, and the results could not be entirely extrapolated to empathic interaction in computer-mediated communication in general. And, of course, the researcher would have to deal with the learning effect problem. One of the suggestions for the experiment design then would be to make the first "pre-session" with just several minutes of conversation while using the app: this would to some extent help with the novelty problem. Yet, the learning effect still is an important factor, so randomizing the order of the experimental conditions and increasing sample size would be required.

Another limitation is related to the explicit valence-arousal scale I used for the study. Like in any self-report research, it is difficult to distinguish participants' sincere emotional state from their intentions to show the 'expected' emotional behavior. Hence, all participants were instructed to state what they feel and not to try to follow somebody else's expectations. I emphasized there were no "right" or "wrong" pictures in the app and the images served solely to reflect (and express) emotions.

Obtained data and the visualizations in Figure 25 and Figure 26 in Qualitative observation part show that points of coordinates in the valence-arousal scale in most cases correspond with participants' self-reported emotional state.

In general, there is a specific view on emotional interaction in this work: I do not aim to measure objective parameters (e.g., heart rate, muscle movements, etc.) but subjective impressions of participants' feelings during interpersonal communication. Hence, if the participant feels that she approaches the emotional state of her partner (or feels something completely different) this subjective feeling is more important than the cortisol level. For more details, see Interaction Process & Emotions chapter.

4.3 POTENTIAL IMPACT AND FUTURE RESEARCH

There are multiple potential impacts from this study. The research has proved that using the app with the valence-arousal scale is justified: the participants' points of coordinates are correlated with their reported subjective emotional states. This means the method is relevant for investigating empathic interpersonal communication through technologies.

The created "Emotional map" (Affective Picture database) for reflecting and expressing emotional state on the basis of weather pictures is another important impact. This map can be used for any other studies related to empathic interaction in computer-mediated communication. The map will be available online for free use in the nearest future.

Several findings and insights (see Main Findings and Insights) may serve as a starting point for future research topics: empathic interaction nature (based on the difference between partners A and B), different empathic strategies/trajectories while discussing sensitive topics ("mood paths"), etc. Besides, I expect that replicating the study with large sample size and/or pairs of good friends and applying within-group design will help to prove the existing hypothesis of this study.

4.4 Summary and Conclusion

In this study, my aim was to investigate how we can enhance the empathic experience in computer-mediated communication. Based on social interaction views by Hanne De Jaegher and the affective interaction approach in HCI by Kirsten Boehner, I developed a hypothesis that interactive emotional cues in communications in technologies will enhance the empathic experience (by increasing the emotional approachment of the partners). To test this, I developed an experiment where two groups of participants (test and control) had to communicate through an audio-conference tool while using an app specially developed for this study. The idea of the app relies on Russell's Circumplex model of emotions (which states that all emotions can be represented in two-dimensioned valence-arousal scale) and Gärdenfors' theory of Conceptual spaces (which sets the idea that similar objects are "situated" closer to each other in n-dimensional space). Test and control groups were distinguished depending on the presence of emotional interactive cues in the app which were represented by weather images arranged in the valence-arousal scale. The position of the images was preliminarily validated in the preparational study, and now this system of coordinates forms a so-called "Emotional Map" (Affective Picture database). After conducting 28 experimental sessions with 56 participants (14 sessions per group), I have analyzed the differences between the groups in their behavior (by analyzing the starting and final points of emotional states / coordinates in the scale) and self-reports (questionnaire and joint reflection).

Due to big individual differences and the small sample size, the results did not prove the declared hypothesis (although, they did not reject it as well). A significant difference in approachment has been recorded, however, due to a significant difference in the starting points, not the final points among control and test group participants. This may signify a strong noise in the data (or an important external factor that has not been uncovered by the researcher).

Still, this research achieved two other goals of the study: created "Emotional Map" - the new Affective Picture database on the valence-arousal scale (that represents the internal emotional state as opposed to the IAPS, for instance, that captured emotional responses from the pictures), and a proved design method of studying empathic interactions in computer-mediated communications. Both approaches in this design method are important for this research. The qualitative self-reports serve as a bridge to quantitative analysis (coordinates and questionnaire) in the study. They validate findings and help in interpretation. And vice versa, the numbers that were collected in the system of coordinates, provide the ground for the qualitative observation. Together, these research approaches represent a comprehensive study of emotional communications.

The main goal of this thesis was to help people to communicate with the assistance of technologies - to increase the quality of their communication. I aimed to do this through enhancing users' empathic experience, which, on its part, will benefit humans' relationships and well-being. I hope the results from this research will find their applications in both computer-mediated communications and emotions studies field.

5 REFERENCES

Andreeva, G. M. (1988). *Social Psychology: Textbook for students of higher educational institutions studying in the specialty "Psychology"* (2nd ed.). Moscow University Publishing House.

Banaee, H., & Loutfi, A. (2014). *Using Conceptual Spaces to Model Domain Knowledge in Data-to-Text Systems*. https://doi.org/10.3115/v1/W14-4403

Bickmore, T. W., Fernando, R., Ring, L., & Schulman, D. (2010). Empathic Touch by Relational Agents. *IEEE Transactions on Affective Computing*, *1*(1), 60–71. https://doi.org/10.1109/T-AFFC.2010.4

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, *65*(4), 275–291. https://doi.org/10.1016/j.ijhcs.2006.11.016

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 379–390.

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

Bradley, M. M., & Lang, P. J. (2017). International Affective Picture System. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–4). Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_42-1

Broadband subscriptions per 100 people. (n.d.). Our World in Data. Retrieved March 14, 2021, from https://ourworldindata.org/grapher/broadband-penetration-by-country

Bruns Alonso, M., Hummels, C. C. M., Keyson, D. V., & Hekkert, P. P. M. (2013). Measuring and adapting behavior during product interaction to influence affect. *Personal and Ubiquitous Computing*, *17*(1), 81–91. https://doi.org/10.1007/s00779-011-0472-3

Colman, A. M. (2015). *A Dictionary of Psychology*. Oxford University Press.

Coplan, A., & Goldie, P. (Eds.). (2011). *Empathy: Philosophical and psychological perspectives*. Oxford University Press.

Cozza, V., Fenu, G., Scateni, R., & Soro, A. (2011). *Walk, Look and Smell Through*. ACM SIGCHI Italian Chapter (CHItaly 2011). https://api.semanticscholar.org/CorpusID:147567937

Daily hours spent with digital media. (n.d.). Our World in Data. Retrieved March 14, 2021, from https://ourworldindata.org/grapher/daily-hours-spent-with-digital-media-per-adult-user

Damasio, A. R. (1996). *Descartes' error: Emotion, reason and the human brain*. Papermac.

De Jaegher, H. (2015, April 27). *Talk Introduction to Participatory Sense-making at International Convention of Psychological Science 2015*. Participatory Sense-Making. https://hannedejaegher.net/2015/04/27/video-introduction-to-participatory-sense-making/

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507. https://doi.org/10.1007/s11097-007-9076-9

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

El Kaliouby, R. A. (2005). *Mind-reading machines: Automated inference of complex mental states.* University of Cambridge, Computer Laboratory 2005-07.

Fogel, A., & Garvey, A. (2007). Alive communication. *INFBEH Infant Behavior and Development*, *30*(2), 251–257.

Gärdenfors, P. (1996). Conceptual Spaces as a Basis for Cognitive Semantics. In A. Clark, J. Ezquerro, & J. M. Larrazabal (Eds.), *Philosophy and Cognitive Science: Categories, Consciousness, and Reasoning* (pp. 159–180). Springer Netherlands. https://doi.org/10.1007/978-94-015-8731-0_8

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. A Bradford Book. http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=32643&site=ehost-live

Gärdenfors, P. (2011). Semantics Based on Conceptual Spaces. In M. Banerjee & A. Seth (Eds.), *Logic and Its Applications* (Vol. 6521, pp. 1–11). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-18026-2_1

Gosain, D., & Sajwan, M. (2014). *Aroma Tells a Thousand Pictures: Digital Scent Technology a New Chapter in IT Industry*. Undefined. /paper/Aroma-Tells-a-Thousand-Pictures%3A-Digital-Scent-a-in-Gosain-Sajwan/66113ffa4d86e4158202ce6e85d470c730dda778

Guo, J., Hou, Y. T.-Y., Mueller, H., Tang, K., & Fussell, S. R. (2019). As If I Am There: A New Video Chat Interface Design for Richer Contextual Awareness. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3312759

Hancock, P. A., Pepe, A. A., & Murphy, L. L. (2005). Hedonomics: The Power of Positive and Pleasurable Ergonomics. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *13*(1), 8–14.

Helander, M. G., Tham, M. P., & International Conference on Affective Human Factors Design (Eds.). (2003). *Hedonomics—Affective human factors design*. Taylor & Francis.

Hume, D. (1739). *A treatise of human nature; being an attempt to introduce the experimental method of reasoning into moral subjects.*

Jeon, M. (2017). *Emotions and affect in human factors and human-computer interaction*. http://www.sciencedirect.com/science/book/9780128018514

Kim, S., Junuzovic, S., & Inkpen, K. (2014). The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. *Proceedings of the 18th International Conference on Supporting Group Work - GROUP '14*, 167–177. https://doi.org/10.1145/2660398.2660409

Koeze, E., & Popper, N. (2020, April 7). The Virus Changed the Way We Internet. *The New York Times*. https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. John Wiley & Sons.

Lee, J.-S., Kwon, H.-M., Shin, D.-Y., & Song, J.-B. (1999). Force feedback using sensibility ergonomics theory in teleoperation system. *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No.99CH36289)*, *1*, 597–602 vol.1. https://doi.org/10.1109/IROS.1999.813069

Maibom, H. (2017). *The Routledge Handbook of Philosophy of Empathy* (H. Maibom, Ed.; 1st ed.). Routledge. https://doi.org/10.4324/9781315282015

Maturana, H. R., & Varela, F. J. (2008). *The tree of knowledge: The biological roots of human understanding*. Shambhala.

Mehrabian, A. (1980). *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies*. Cambridge : Oelgeschlager, Gunn & Hain. http://archive.org/details/basicdimensionsf0000mehr

Mobile cellular subscriptions by country. (n.d.). Our World in Data. Retrieved March 14, 2021, from https://ourworldindata.org/grapher/mobile-cellular-subscriptions-by-country

Nagamachi, M. (1995). Kansei Engineering: A new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics International Journal of Industrial Ergonomics*, *15*(1), 3–11.

Naotaka, F., Sayaka, H., & Atsushi, I. (2007). Dynamic Social Adaptation of Motion-Related Neurons in Primate Parietal Cortex. *PLOS ONE*, *4*.

Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). *Improving automotive safety by pairing driver emotion and car voice emotion*.

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. https://central.bac-lac.gc.ca/.item?id=TC-BVAU-2863&op=pdf&app=Library

Ortiz-Ospina, E., & Roser, M. (2020). Loneliness and Social Connections. *Our World in Data*. https://ourworldindata.org/social-connections-and-loneliness

Percentage of individuals using the internet by country. (n.d.). Our World in Data. Retrieved March 14, 2021, from https://ourworldindata.org/grapher/percentage-of-individuals-using-the-internet-by-country-2000-2014

Picard, R.W. (1997). *Affective computing*. MIT Press. http://www.books24x7.com/marc.asp?bookid=1283

Picard, R.W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(10), 1175–1191. https://doi.org/10.1109/34.954607

Pixabay License. (n.d.). Retrieved January 30, 2021, from https://pixabay.com/service/license/

Plutchik, R., & Kellerman, H. (1989). *The Measurement of emotions*. http://site.ebrary.com/id/10989247

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *HUMOV Human Movement Science*, *26*(6), 867–891.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Shapiro, L. A. (2014). *The Routledge handbook of embodied cognition*. https://www.taylorfrancis.com/books/9781317688662

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and Coordinative Structures. *TOPS Topics in Cognitive Science*, *1*(2), 305–319.

Soegaard, M., & Dam, R. F. (2013). *Encyclopedia of Human-Computer Interaction*.

Stein, E., & Stein, W. J. H. (1989). *On the problem of empathy*. ICS Publications.

Stueber, K. R. (2006). *Rediscovering empathy: Agency, folk psychology, and the human sciences*. MIT Press.

Sundström, P., Ståhl, A., & Höök, K. (2005). *eMoto—Affectively Involving both Body and Mind*. http://soda.swedishict.se/148/1/sp393-sundstrom.pdf

The rise of social media. (n.d.). Our World in Data. Retrieved March 14, 2021, from https://ourworldindata.org/rise-of-social-media

Titchener, E. B. (1909). *Lectures on the experimental psychology of the thought-processes*. New York, The Macmillan company. http://archive.org/details/cu31924024573176

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. Journal of Child Psychology and Psychiatry and Allied Disciplines, *42*, 3–48.

Varela, F. J., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Currents in modern biology, 5(4), 187–196. https://doi.org/10.1016/0303-2647(74)90031-8*

Varela, F. J., Rosch, E., & Thompson, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT.

Walla, P., Nesbitt, K., Blackmore, K., Hookham, G., & Kay-Lambkin, F. (2015). *Using the Startle Eye-Blink to Measure Affect in Players* (pp. 401–434). https://doi.org/10.1007/978-3-319-05834-4_18

Watson D & Tellegen A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235.

Xu, B. (2017). *Designing Computer Mediated Communication through Alternatives: Digital Ephemerality, Content De-emphasis and Their Implications*. https://doi.org/10.7298/X4B56GX8

Zahavi, D. (2001). Beyond Empathy: Phenomenological Approaches to Intersubjectivity. *Journal of Consciousness Studies*, *8*, 151–168.

6 APPENDICES

6.1 ABSTRACT (ENGLISH)

Objectives: The study's primary purpose was to research how to enhance the empathic experience in computer-mediated communication (CMC) - the channel people tend to use more often these days. According to the hypothesis, it can be achieved by adding extra interactive emotional cues in audio conversation. The additional goals were to test the plausibility of the new method of emotion research and, in case of success, to contribute to the research community with the new Affective Picture database.

Methods: To test the hypothesis, a special app was developed. It represented emotional states by weather pictures situated on a valence-arousal (v-a) scale. This v-a scale – Emotional map – is based on a preliminary conducted survey (n=317). In a two-group design experiment, empathic experience of the group that exchanged interactive emotional cues (weather images) was compared with the control group without extra interactive signal. The comparison was made on parameters: emotional closeness (tested method: the distance between the participants' points of coordinates before and after conversation), and subjective experience of connectedness, comfort, and understanding each other's emotions (accompanying methods: Likert-scale survey and joint reflection).

Results: The comparison of emotional closeness showed a significantly greater change of distance after the conversation in the group with interactive cues ($p = .03$). However, it happened due to a significantly bigger distance in the beginning, hence, the replication with larger sample size is required to eliminate the noise. Yet, analysis of "mood paths" (participants' path directions from the start to the end) showed several different behavior strategies. Visual overview indicated some responsive behavior supplanted by summarizing statistics, hence this approach is worth to be investigated further. Additionally, a significant difference was found in partners' behavior depending on one's initial emotional state: bad mood agents tend to increase their valence level more than their positive partners tend to decrease theirs. This finding may be useful for studying empathic interactions in general, not only in CMC. Comparison of the numeric data (emotional state expressed by points of coordinates) with the qualitative observations (participants' self-reported emotional state) validated the applied research method. These results proved the plausibility of the method and Affective Picture database for use in emotional research.

Conclusion: Besides contributing to digital empathic experience - oftentimes impoverished - this research may benefit a general understanding of emotions and empathy. Further studies are required to examine the influence of interactive emotional cues on empathic experience. Still, the main research findings can serve as a strong base for future research in these fields.

Keywords: Affect, emotions, empathy, interpersonal communication, computer-mediated communication, HCI, affective interaction, social interaction

6.2 ABSTRAKT (DEUTSCH)

Ziele: Der Hauptzweck der Studie bestand darin, zu untersuchen, wie die Erfahrung von Empathie in der computervermittelten Kommunikation (CVK) verbessert werden kann - die Art der Kommunikation, die Menschen heutzutage also immer häufiger nutzen. Die Hypothese ist, dass dies durch Hinzufügen zusätzlicher interaktiver emotionaler Signale in Audiokonversationen erreicht werden kann. Das zusätzliche Ziel bestand darin, die Plausibilität der neuen Methode der Emotionsforschung zu testen und im Erfolgsfall mit der neuen "Affective Picture"-Datenbank einen Beitrag zur Forschungsgemeinschaft zu leisten.

Methoden: Um die Hypothese zu testen, wurde eine spezielle App entwickelt. Diese stellte emotionale Zustände durch Wetterbilder dar, die entsprechend auf einer Valenz-Erregungs-Skala (v-a) eingeordnet waren. Diese Emotionskarte der VA-Skala basiert auf einer vorläufigen Umfrage (n = 317). In einem Zwei-Gruppen-Design-Experiment wurde die empathische Erfahrung der Gruppe, die interaktive emotionale Hinweise (Wetterbilder) austauschte, mit der Kontrollgruppe ohne zusätzliches interaktives Signal verglichen. Der Vergleich wurde anhand von den Parametern emotionale Nähe (getestete Methode: Abstand zwischen den Koordinatenpunkten der Teilnehmer vor und nach dem Gespräch) und subjektive Erfahrung von Verbundenheit, Komfort und gegenseitigem Verständnis der Emotionen (begleitende Methoden: Likert-Skala Umfrage und gemeinsame Reflexion) durchgeführt.

Ergebnisse: Der Vergleich der emotionalen Nähe zeigte eine signifikant größere Änderung der Distanz nach dem Gespräch in der Gruppe mit interaktiven Hinweisen ($p = 0{,}03$). Dies geschah jedoch aufgrund einer erheblich größeren Entfernung zu Beginn, weshalb die Replikation mit größerer Stichprobe erforderlich ist, um das Rauschen zu beseitigen. Die Analyse der „Stimmungspfade" (Pfadrichtungen der emotionalen Zustände der Teilnehmer von Anfang bis Ende) deutete auf verschiedene Verhaltensstrategien hin. Die visuelle Übersicht zeigte ein gewisses Reaktionsverhalten an, das durch die Zusammenfassung von Statistiken ersetzt wurde. Daher sollte dieser Ansatz weiter untersucht werden. Darüber hinaus wurde ein signifikanter Unterschied im Verhalten der Partner in Abhängigkeit vom anfänglichen emotionalen Zustand festgestellt: Personen mit negativer Valenz neigen dazu, ihre Werte stärker zu erhöhen als ihre positiv gestimmten Partner, ihre zu verringern. Diese Einsicht kann nützlich sein, um empathische Interaktionen im Allgemeinen zu untersuchen, nicht nur bei CVK. Der Vergleich der numerischen Daten (emotionaler Zustand ausgedrückt durch Koordinatenpunkte) mit den qualitativen Beobachtungen (Selbstbericht über emotionalen Zustand durch Teilnehmer) bestätigte die angewandte Forschungsmethode. Diese Ergebnisse bewiesen die Plausibilität der Methode und der "Affective Picture"-Datenbank für die Verwendung in der Emotionsforschung.

Schlussfolgerung: Neben dem Beitrag zu, der oft digital reduzierten empathischen Erfahrung kann, diese Forschung ein allgemeines Verständnis von Emotionen und Empathie fördern. Weitere Studien sind erforderlich, um den Einfluss interaktiver emotionaler Signale auf das empathische Erleben zu untersuchen. Dennoch können die Forschungsergebnisse als starke Grundlage für die zukünftige Forschung in diesen Bereichen dienen.

Schlüsselwörter: Affekt, Emotionen, Empathie, zwischenmenschliche Kommunikation, computervermittelte Kommunikation, HCI, affektive Interaktion, soziale Interaktion