Digital Humanity:
The Temporal and Semantic Structure of Dynamic Conversational Facial Expressions

Von der Fakultät 1 - MINT - Mathematik, Informatik, Physik,
Elektro- und Informationstechnik
der Brandenburgischen Technischen Universität Cottbus–Senftenberg
genehmigte Dissertation
zur Erlangung des akademischen Grades eines

Dr.-Ing.

vorgelegt von

Susana Castillo Alejandre

geboren am 21.03.1980 in Zaragoza, Spain

Vorsitzender: Prof. Dr.-Ing. habil. Michael Hübner
Gutachter: Prof. Dr. habil. Douglas W. Cunningham
Gutachter: Prof. Dr. rer. nat. habil. Michael Breuß
Gutachter: Prof. Dr.-Ing. Marcus Magnor
Tag der mündlichen Prüfung: 7.6.2019

ABSTRACT

This thesis focuses on establishing a – theoretically founded and empirically derived – novel methodological pipeline to provide Embodied Conversational Agents (ECAs) with natural facial expressions and desired personality traits.

After giving an overview on the content of this thesis, we dedicate its second part to derive the Semantic Space (SSp) for facial expressions, finding that the same space is used for expression words, expression videos, and motion-capture-based point-cloud animations. The process involved the creation of a new facial expression database using Motion Capture (MoCap) technology. Our technique can be used to empirically map specific motion trajectories (including their frequency-decomposition) onto specific perceptual attributes, allowing the targeted creation of novel animations with the desired perceptual traits, as exemplified in the third part of this thesis. Before addressing our final conclusions and, on the grounds that the systematic differences between individuals while performing the same facial expressions are related to their personality, we devote the fourth part of this thesis to the study of the mapping between personality and expressive facial motions.

MEASURABLE CONTRIBUTIONS OF THE AUTHOR

The following contributions, detailed in Section 1.3, were the result of this thesis :

- 3 JCR-indexed journal publications [1, 2, 3]
- 6 peer-reviewed conference publications [4, 5, 6, 7, 8, 9]
- Participation in 2 research projects
- Reviewer for 5 journals and 2 international conferences
- Poster Chair for one international conference
- 5 invited talks
- Teaching Assistant during 12 semesters for 3 different lectures
- Lecturer during 4 semesters for 3 different courses
- 20+ co-advised BSc and MSc theses

ZUSAMMENFASSUNG

Diese Arbeit konzentriert sich darauf eine – theoretisch fundierte und empirisch belegte – neue methodologische Pipeline bereitzustellen, um Embodied Conversational Agents (ECAs) mit natürlichen Gesichtsausdrücken und gewünschten Persönlichkeitseigenschaften auszustatten.

Nachdem ein Überblick über den Inhalt der Arbeit gegeben wurde, wird sich der zweite Teil der Dissertation mit der Herleitung des Semantic Space (SSp) für Gesichtsausdrücke beschäftigen und dem Belegen das selbiger SSp sowohl für die Wörter, die zum Beschreiben der gegebenen Gesichtsausdrücke genutzt werden, als auch für Kameravideos und für Motion-Capture basierte Punktwolken-Animationen genutzt werden kann. Dafür musste eine neue Ausdrucksdatenbank mit Hilfe der Motion Capture (MoCap) Technologie angefertigt werden. Unsere Technik erlaubt es bestimmte Bewegungskurven (inklusive deren Frequenz) auf spezifische wahrgenommene Attribute abzubilden und damit gezielt neue Animation mit gewünschten Persönlichkeitseigenschaften zu erschaffen, dies wird im dritten Teil der Dissertation veranschaulicht. Bevor wir abschließende Schlussfolgerungen ziehen, widmet sich der vierte Teil der Dissertation, basierend auf der Grundlage, dass unterschiedliche Individuen gleiche Gesichtsausdrücke auf Basis ihrer Persönlichkeit anders darstellen, einer Studie, welche eine Abbildung von Persönlichkeit und ausdrucksvollen Gesichtsbewegungen vorsieht.

ACKNOWLEDGMENTS

There are many people (too many to mention) that have had an impact on my life and work since I started walking this path. I would like to thank them all and some in particular.

First, I would like to thank my advisor, Douglas Cunningham, not only for guiding me through all these years and making me a better scientist, but also for becoming a friend. I will always be thankful to him and his family for the warm welcome they gave me to Germany and their help to make Cottbus my home.

I am more than grateful for having had Katharina Legde as my officemate, it has been a pleasure to work with her, and I am lucky to have counted with her support all along these years in every aspect of my life. I would also like to thank Philipp Hahn for all the good times together even during the hectic deadlines, Martin Schorradt for being both my metal buddy and the student who will always make me proud, and Mikhail Ashkerov for contributing to the multiculturality of the group.

I would also like to show my gratitude to Stephan Guthe, Christian Wallraven, and Michael Breuß for all the insightful scientific conversations we had and all their really appreciated ideas.

Special thanks go to all present and former colleagues of the Computer Graphics Lab at BTU Cottbus-Senftenberg. Specially I would like to thank Christian Winger for all the time an effort he put on helping me in my beginnings in Germany, not only academically but socially, and to Mike Schönwiese for the selflessly dedicated time to help when was needed.

I could not but thank Diego Gutierrez for opening this world to me, and all my former colleagues from the Graphics & Imaging Lab in Zaragoza, without whom I will probably not be here today. Special mentions to Jorge Jimenez, Adrian Jarabo, Jose Ignacio

Echevarria, Carlos Aliaga, Jorge Lopez-Moreno, Adolfo Muñoz, Elena Garces, and Belen Masia.

An extended thanks goes to my Spanish friends for letting me know that I am not forgotten, and to my German ones, especially to Franzi and Thomas, for helping me feeling at home in a new country.

Last but not least, I need to thank my family for letting me go to pursue my future abroad and making sure that I feel like I never left every time I go back to Spain. Special thanks to my parents and my brother for being a pillar in my life and have made this dream of mine come true, and being always by my side, even when it was not always easy.

This research was partially funded by the German Research Council, under Grant CU 150/2-1.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ECA Embodied Conversational Agent

DB Data Base

SSp Semantic Space

PCA Principal Component Analysis

SSq Sum of Squared Errors

SEM Standard Error of the Mean

MoCap Motion Capture

SAL Sensitive Artificial Listener

FFMRF Five-Factor Model Rating Form

MPI Max Planck Institute

BTU Brandenburg Technical University

Part I

Introduction & Overview

1 INTRODUCTION

Communication is inherent to life. All species communicate with their conspecifics and, sometimes, with specimens of other species. In the case of humans, this need to communicate is unavoidable; we cannot not communicate. We sure establish some kinds of information exchange with other species but, as it is to be expected, we treat humans differently because we share mental and emotional states with other members of our kind. Human communities are based on the sharing of certain values and we have some social protocols we need to fulfill, and rules we need to follow, not to be outcasts from the society. There is a full field which studies all the patterns that allow us to behave and understand other members of our society, the science of Social Intelligence. To show these behavioral patterns – and to perceive them from our interlocutors – we focus on our actions.

It is just natural to us, then, to treat human-looking entities as we would treat our fellow men. Embodied Conversational Agents (ECAs) are virtual characters who are able to interact with human beings by interpreting and producing multimodal communicative behavior. If these ECAs are going to be treated as humans, we need to design them to fulfill the expectations people put on them. This way we will make them more believable: the ECAs will give the impression to think and behave like us, even when this is just an illusion and there is no real socio-emotional intelligence behind these virtual agents. Of course, there is a lot of variance in these expectations among communities. What we consider to be an appropriate answer or expression of our emotions is highly influenced by our culture. Therefore, in this thesis, we pay special attention to the cultural dependencies when designing our experiments and deriving our conclusions,

as we are aware that what can be true for our specific scenarios does not necessarily need to be true for all humanity. Nevertheless, and despite the cultural dependencies on both the design of an specific ECA and its target audience, we can assume that the core requirements to make an ECA human-like are universal. One might consider that the ECA should fulfill three main requisites: to have an identity, to look human, and to behave like a human. In this thesis, we will focus on this last aspect, the socio-emotional information carried by the actions of the ECA during communication.

It has been previously shown that when interacting with an interlocutor, it is more important what our face expresses than what our words actually say. An impressive 55% of the affective meaning is transferred via facial expressions, 38% through the help of prosody (e.g., the acoustic modulations related to speech melody, speed, and intensity) and only 7% is conveyed via pure spoken text [11]. This predominance is reinforced by the fact that non-verbal aspects are given more weight if there are discrepancies between non-verbal information and the words [12]. Furthermore, we can decide to stay silent but we can not avoid producing non-verbal signals, even when we would rather not [13]. Consequently, in order to narrow the spectrum of actions to be studied and replicated without weakening the impact of our research, we decided to focus on conversational facial expressions.

On the one hand, the Computer Graphics community has come a long way in the quest to make virtual agents look more and more realistic. Nowadays, we have a broad variety of sophisticated animation techniques at our disposal that enable the almost perfect replication of reality. The Computer Graphics community is close to mastering the "how to" of animation, nevertheless, the "what to" animate still poses some challenges. For example, either via manual animation (i.e. through the hands of an artist), or via the technique of Motion Capture, we can mimic/transfer the motion of a real person to a virtual agent. But, in both scenarios, who decides the specific set of facial deformations to be transferred that will convey the desired emotional state or reaction, is always a human; either the artist, or the recorded actor. Hence, to be able to *automatically* generate facial expressions while being sure that their conveyed meaning is the intended one, we need to establish a mapping between facial deformations and meaning.

On the other hand, within Psychology, the study of the structure of meaning and its relationship to facial deformations is well established. The research conducted in this discipline tends to be based on static stimuli and the study of recognition, intensity and

sincerity rates of the stimuli. Facial expressions do more than simply convey an emotion [12, 14]. Among others, they provide information on how we are feeling, what we think of the other people in the conversation, our state of health, about our social skills in general, and the relationship(s) between us and the others in the conversation. Moreover, it has been proven that facial expressions provide emphasis [15] on part of the message, modify the information provided by another channel [16], and control the flow of the conversation [17].

This methodology at best establishes only a qualitative mapping between parts Socio-Emotional meaning and expressions. It captures only a fraction of the information communicated by expressions, does not establish a quantitative one-to-one mapping between the complexity of a metric Socio-Emotional Space and specific facial deformation. We argue that this methodology is not sufficient for the direct control and/or production of animations.

Given the complementary expertise, it is clear that increasing the synergy between the two disciplines should benefit both. Harnessing these aspects of non-verbal communication can lend virtual agents greater depth and realism, by giving them the ability to actually produce social information. With the purpose to be able to provide the virtual agent with the appropriate capacity to perform natural facial expressions, a more comprehensive, metric mapping between actions and Socio-Emotional Space is required.

Since our expressions are the reflection of our inner mental state, they do not solely reflect our reaction to a particular stimulus. We hypothesize that they get colored by our mood and personality. In this thesis, we make a novel proposal; *the duration of a mental or emotional state is directly related to the permanency of it, and therefore to the depth of the information it provides about an individual.* Punctual ephemeral reactions (experienced for a short period, from a few seconds to a couple of minutes) are fleeting, and represent our current state of mind triggered by recent stimuli – e.g. we laugh when we are told an amusing joke. When the duration of a particular state of mind is prolonged for dozens of minutes, a few hours or even a couple of days, it represent something more stable, a temporary disposition to reality, generally refer to as our "mood" – e.g. after being told we lost a competition, we are in low spirits for the next hours. Finally, longer, more permanent reactions relate more to our underlying disposition or personality – e.g. our tendency to always look at the bright side of life, makes us an optimistic person. Note that these reactions are not mutually exclusive and, in fact, they co-occur, so when

we show a punctual reaction to a stimulus, it always comes embellished by these other emotional levels. Figure 1.1 gives a graphical representation of this concept.

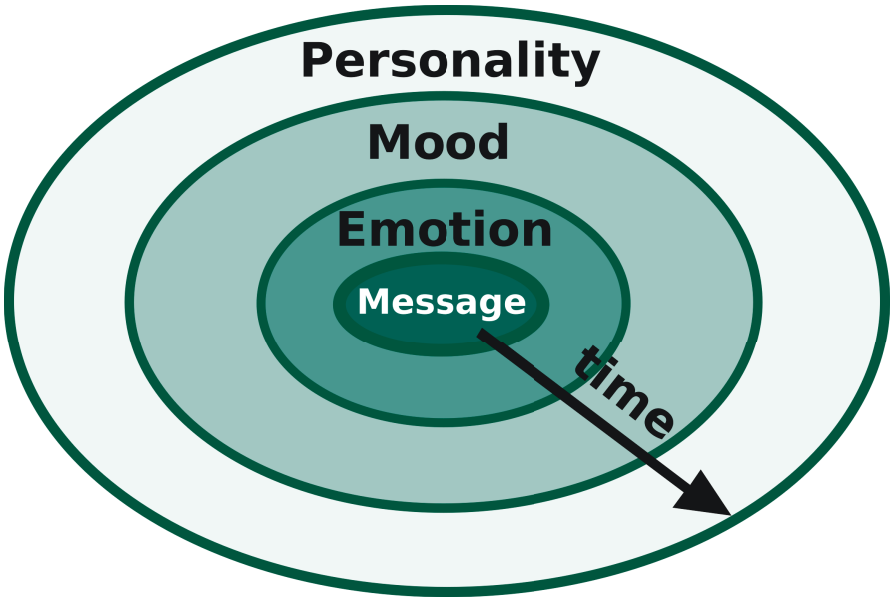

Figure 1.1: Visualization of the proposed relation between the duration of a mental or emotional state and its permanency.

For example, imagine we are having a conversation with a coworker that we know well, and know to posses a cheerful, positive personality. If we present this individual with some good news and they smile, but not quite as full-heartedly as usual; they stay engaged in the conversation and their expressions match the content of the discussion, but are shorter and sharper than normal for them; and their negative expressions are more frequent and stronger than what we are used to see in that person, we will infer that our colleague is in a "bad mood" that day. Reaching this conclusion will probably also affect the way we will perceive the punctual reactions of our colleague from that moment on. For example, by being able to distinguish which of our comments genuinely annoy them and which do not.

Naturally, we can also see these three layers playing a role when we first meet someone. For example, if we are a third person meeting for the first time the aforementioned colleague that same day, we could associate their harsh negative reaction to a comment we make, to the comment itself. If we keep on experiencing this kind of angry vibe in the way that person expresses him or herself, we would start thinking that either that

person has bad temper or is having a bad day. Posterior encounters and observations of that same person along an extended period of time, would allow us to distinguish between the two options. If the expressions of the individual always follow the same pattern, we will conclude that angriness is, indeed, part of their personality.

It seems to be clear then, that all this complexity reaches somehow our interlocutor during the interaction. Neglecting these intricacies of expressions when animating a virtual agent and failing to reflect them, will most likely lead us to undesired outcomes, as we can not forget that people have inherently learned to always interpret non-verbal signals in human-like communication.

To capture the "Gestalt" perception of an ECA, this thesis proposes the use of traditional psychological methods to recover the vector space – the Semantic Space (SSp) – underlying facial expressions while including their full range of nuances.

1.1 Goal & Pipeline

The overarching aim of this thesis is to understand the structure, motion, and meaning of facial communication and to use that with virtual characters – in particular ECAs – in order not only to ensure that the agents can convey the meaning intended, but also to communicate with their interlocutors in a natural, human-like way. That is, the goal aims at an understanding of the semantic nature of expressions and how that maps to the physical structure, so that synthetic facial expressions are as natural-looking as possible. To achieve that goal, we need to find a characterization of the dynamic structure of facial expressions on two different levels.

The first level focuses on obtaining the higher-level semantic structure underlying the facial expression space. For example, surprise, happiness, and disgust might seem absolutely different but, in fact, the three of can be compared and classified according to many criteria, e.g. their valence (neutral, positive, or negative), their predictability (unexpected or predictable), and their energetic level (high, mid or low intensities). As it has been previously shown [18], emotions share some common features and, therefore they can be represented in a vector space, our Semantic Space (for more details, please refer to Chapter 2).

The second level focuses on formally describing the physical structure of facial expressions (see Chapter 3). In particular, this level describes the sources of information that humans use to perceive different aspects of a facial expression. These physical descriptions include determining which movements are the ones that are necessary and sufficient for a given expression to be perceived.

The link of the first and second levels would allow one to map the Semantic Space to the physical motions in order to create naturalistic facial expressions (see Chapter 4).

Nevertheless, to provide virtual agents with full human-like communicative capabilities, natural facial expressions are not enough (see Chapter 5). We would like to impregnate all emotions with coherent subtle behavioural deviations that can be perceived as series of personality traits which will result in the classification of the virtual character as a particular kind of individual (e.g. an aggressive character or a cheerful one). Thus, this thesis also explores how to provide the agent with a personality that can be appreciated by the interlocutor, analyzing the implicit indication of the mapping between the Semantic Space for facial expressions and the Personality Space (see Chapter 6).

The final conceptual architecture of this thesis, for which we will give more details in the following, is illustrated in Figure 1.2.

Please note, that deriving the Semantic Space indicated in the architecture is a core element of our work. To be able to obtain it, a consistent methodology is used throughout this thesis. Three main tools are consistently employed to gather the vector spaces determining the structure of our visual stimuli: Likert scales [19], Semantic Differential Task [20], and Factor Analysis (an extended version of Principal Component Analysis (PCA)). The core of the experiments to derive the vector space, that establishes the connection between meaning and visual information (the SSp), consists of asking participants to rate the visual stimuli along several Likert scales, with each end of a scale being anchored by a pair of opposites. These psychometric scales are known to be able to measure, through equally distant choices, the level of agreement/disagreement with a given assessment [19]. Once the ratings for all stimuli are gathered, their covariation is examined using factor analysis to extract the underlying structure. For more on these tools, please see the work of Cunningham and Wallraven [21].

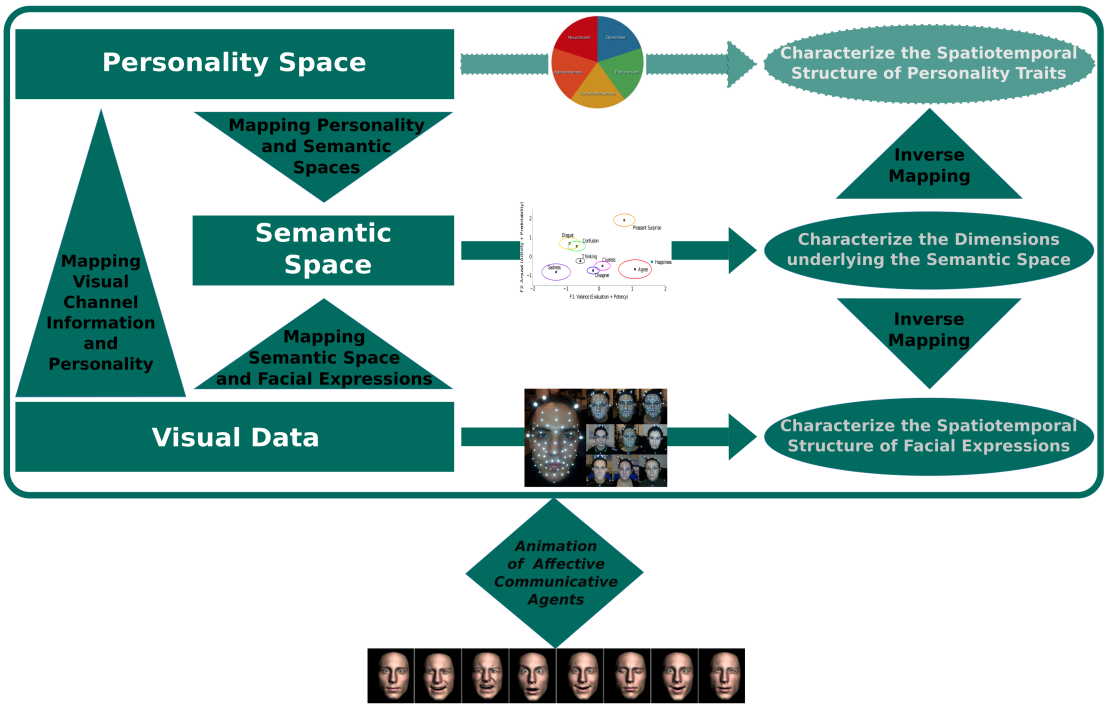

Figure 1.2: Extended conceptual architecture of this thesis.

1.1.1 Finding the Semantic Space for Facial Communication

To better understand facial communication, including the expression of emotions, we needed to recover the metric structure of the Semantic Space. To do this, we perform the experiments detailed in Chapter 2, following an extension of Osgood's semantic differential technique [20] including some of Fontaine's adaptation to emotional words [18]. This technique allows one to recover the underlying structure of a stimulus set by first rating them along a series of scales, and then using either PCA or its extended version, factor analysis, to examine the covariation of the data. One interesting side-effect of using multiple people in the recordings is that the difference between people for a given set of expressions revealed some elements of individual personality. That is, although all the recorded individuals had a remarkable consistency in how they performed a given expression as well as how that expression was perceived, there were systematic differences between people (across expression), that seem to be due to underlying personality and mood differences.

The results of this part of the thesis (in particular, Part II) tell us how similar the different versions of a given expression are (and how a given person differs from others, across expressions). It also shows the important dimensions of facial expressions (such as evaluation, power, activity). Since this is a metric space (the distances between emotions in the space can be calculated in a straightforward manner and are externally valid), we can then select novel points in the recovered space (where no expressions were recorded) and determine how an expression located there might be perceived using a weighted sum of all (or just nearby) points in the Semantic Space. Furthermore, since we also know which actual videos or MoCap data gave rise to the measured points, a weighted combination of the actual recordings (for example, using morphing, optimal blending, or blend-shape animation) allows us to generate any desired facial expressions corresponding to any point in the space. A practical example for this motion synthesis is given in Part III.

1.1.2 Mapping Semantic Space and Facial Expressions

The next step is to determine which variations in the Semantic Space are caused by which facial movements. Although this has been done to some degree – on a recording by recording basis – in the last step, here we sought in to find commonalities across recordings. That is, we wished to determine which aspects of a given structure's motion are related to personality or mood (and thus unique to that recording or person) and which are related to the actual expression (and thus are in every recording of that expression). This requires, as a first step, exploring the trajectories of the motion tracking points for a given emotion. Since the individual markers are less interesting than the facial regions they lie on, after the motion of individual markers are analyzed, collections of markers are analyzed either by clustering significant areas of the face (mouth, eye-browns, etc.) or in a data-driven manner (by collecting markers that share similar motion parameters such as intensity, duration, peaks, plateaus, and rapidness). The result is a correlation between variations in meaning and the physical changes that are correlated with them. By combining these physical changes using weights determined by distances in the Semantic Space, as described above, we should be able to reproduce any expression in the Semantic Space. That is, we can find a set of coordinates in the Semantic Space (for the desired expression) and then look up the motion parameters that are needed (i.e., which markers to move, in which directions and how fast and intense).

For results and more details on this procedure, please see Part III.

In order to prove that the perception of meaning can be conveyed largely by motion and does not require some other informational channel (such as facial geometry, body posture, voice...), we performed the experiments detailed in Chapter 3. These experiments showed that participants' perception of expressions using videos from real people performing an expression was the same as when the stimuli were merely (moving) point clouds of the facial makers animated with the recorded motion capture data. That is, the location of an expression in the Semantic Space is the same for words, videos, or point-light motion-cloud animations.

1.1.3 Characterize the Spatiotemporal Structure of Facial Expressions

The videos used for Experiment 4 (in Section 3.4 from Chapter 3) contained the raw data gathered from the motion capture database, as the participants were able to detect artifacts in the videos (markers jumping or disappearing and reappearing), these were perceived as such without affecting the perception of the overall expression. Nevertheless, in order to be able to analyze the spatiotemporal information in the perceived expressions, we need to perform a proper computer analysis of the trajectories of the markers. To do so, these artifacts need to be cleaned. We performed a low-pass filtering on the data in order to remove the noise due to incorrect marker information caused by occlusions during the recording, or, in general, degradation of the recorded information. Even if this method is the most commonly used system to remove high frequency data that normally correspond to noise in the recordings, the filter is not aware of the nature of the data that is being filter, removing all high-frequency information that can be present. This is a problem for our purposes, because although part of the high-frequency data is, in fact, noise there are also so-called "micro-movements" (such as a quick smirk, brief twitch of the eyebrows, or a fast wink) which clearly are part of the signal which provide meaningful information. Sadly, a low-pass filter removes both noise and signal, which can significantly alter the meaning of the expression. This is a non-trivial problem to solve, as it requires inserting considerable semantic, context-dependent knowledge into the cleaning algorithm, and no such technique exists. More specifically, before removing part of the motion of a maker, it needs to be determined if that motion is in fact noise or signal (or a combination of the two). This might be possible by examining the motion

of neighboring points (both in the current recording, as well as in other recordings), and with the help of a few additional constraints. Please note that, at the moment this thesis is being written, the development of such a technique is still work in progress, as it is outside the scope of our work. It is clear that there is much more that can be done based on the data and results we provide in this thesis. Although we define a full pipeline, our focus is on being able to derive the mappings between visual information, meaning, and personality. Thus, some elements of the full pipeline, such as semantically-aware motion capture cleaning, remain future work.

1.1.4 Mapping Semantic and Personality Spaces

While recovering the Semantic Space, we made an unexpected finding. First, and not so critical, the perception (as determined by location in Semantic Space) of the recording of an expression is nearly identical to the perception of the word for that expression, on average. For example, when the locations of the expression for pleasantly surprised for all videos (or motion capture recordings) are averaged, it is the same location as the word "Pleasantly Surprised". This further underlines the validity and stability of the Semantic Space. The truly interesting bit is that this is only true for the *average* of all the recordings. Each person has minor but systematic deviations from the average location for each given expression. That is, the set of recordings for a given expression forms an ellipses, the center of which is the location of the (idealized) form of that expression. The distance of a given actor's or actress's recording of a given expression from the average location of that expression can be described as a two-dimensional vector (the derived Semantic Space can be well described with two dimensions). When we look at the set of vectors for a given actor or actress (across expressions), we found a strong correlation that is dependent on the individual (see Sections 2.4.2 and 6.3.1). For example, if a person was located in the positive side of the active dimension and the negative side the evaluative dimension for one expression they tended to have the same deviations for all expressions. That is, they always tended to be seen as more intense or impulsive and negative than average. We think this relation can be an indicator of the personality and mood of the actor or actress, an thereby, an implicit indication of the mapping between the Semantic Space for facial expressions and the Personality Space. We think this finding is worth to be taken into consideration because, being able to replicate this tendency by properly establishing the mentioned mapping between spaces,

will allow us to not only animate an ECA with realistic facial movements, but also to give that ECA a personality that can be appreciated by the interlocutor. If instead of using the generic form of an emotion (i.e., its statistically average location in the Semantic Space) to animate our ECA, we impregnate each expression with coherent deviations from its respective average location based on a series of personality traits, it will result in the classification of the ECA as a particular kind of individual (e.g. an aggressive character or a cheerful one).

Thus, the purpose of Part IV of this thesis is to study the relationship between our Semantic Space and personality. Obviously, towards that goal, first we need to clarify the concept, representation and measurement of personality. Fortunately, Psychology has a very long – and very, very large – history in the study of personality. There are several well-known models that are able to describe the personality of an individual as a collection of ratings among some traits. As with the semantic differentials we used to recover the Semantic Space, these personality ratings can be be compressed into a limited set of dimensions. One of the most successful (and most popular) models is the OCEAN model. It describes personality by measuring a set of traits on five dimensions (Openness, Consciousness, Extroversion, Agreeableness and Neuroticism). For us to be able to make the correlation between the OCEAN space and the Semantic Space, we need to ask participants to rate the perceived personality of the actors while performing Experiment 8 (see Section 6.3, Chapter 6) which is an extended version of Experiment 4 (Section 3.4), by fulfilling a questionnaire that gives us their rating on the OCEAN space. By comparing the position of the actors in this Personality Space and their positions on the Semantic Space we can infer the relation between personality and deviations from the generic expressions. This fusion of both the OCEAN space with the Semantic Space would allow the designers to only need to deal with one set of parameters in order to make the virtual character show one expression at the same time that they give the avatar the desired personality.

Nevertheless, before rushing on trying to apply such a mapping to provide an ECA with a personality, there are several open questions that require our attention. First and most critically, we need to be sure that, as previous research seems to indicate, ECAs are seen as having personalities. Second, we need to assure the validity of standard personality questionnaires to evaluate virtual personalities (the reference corpus used in their design is based on human data). Third, we also need to see the capabilities of such questionnaires to measure personality on uni-modal passive scenarios, i.e. when only

information from one channel, such as the facial movements of an agent, is available and one cannot interact with the ECA. That is, when we only can hear the agent, or when the ECA does not talk but produces facial expressions (as is the case with our expression databases). Finally, assuming virtual agents can have personalities and once the validity of the questionnaires has been established, we can figure out how facial geometry, facial motion, and auditory cues combine to give rise to the overall perception of personality.

Therefore, to provide an initial investigation of these questions, we performed three more experiments (described in Chapter 5) on a state-of-the-art ECA which has four different avatars – two male and two female – each with a distinct, intended personality. In the first experiment for this part, we let people converse freely with the avatar and then rate its personality using a validated short form of standard personality inventory). In this experiment, all perceptual cues were available to the participant, including audio (message and voice) and visual (facial animations and appearance of the avatar). In the second experiment, people were able to talk with the avatar but not to see it (audio only experiment). In the third experiment, participants were presented solely a picture of the avatar (static appearance only). The experiment allow several conclusions to be made. First, virtual characters are seen as having personalities. Second, the proposed personality model (and questionnaire) was appropriate to measure virtual personality. Third, personality is multimodal. Even though personality can be seen on a single channel (audio-only, static picture only), the full personality is only seen when all the cues are present. In other words, the personality seen in a picture or in a audio-only conversation is different than the one perceived when you can see the agent you are conversing with. The combination of the individual channels to make the global personality is not straightforward (and is usually not merely a weighted sum of the ones perceived from the isolated channels). It is also clear, however, that all channels are important to be able to convey the desired personality. Thereby, even if while designing the ECA we need to be aware that different appearance and voice characteristics will affect the impression on the personality given to the interlocutor, and we need to be careful to make them match the ECA's intended personality, having a mapping that establish the relation between facial movements and OCEAN traits can be really helpful for the design of the virtual agent. We explore this relationship explicitly in Chapter 6.

1.2 Summary & Overview

This thesis examines the structure of expressions, both at a semantic level (the perception of meaning) and at a physical level (in the form of facial expressions). A considerable amount of structure at both levels was found, the details of which are given in this thesis, and reported in a series of scientific publications. The two sets of structure were mapped to each other. This not only provides insight into the perception and representation of emotions and other expressions, but also how these are produced. In order to generate the appropriate stimuli as a basis for our research, we recorded a new database of MoCap (and the related real video) of 62 expressions from 10 individuals. During the research process, we had the unexpected and very interesting result that personality seems to also be encoded in specific facial motions, and that personality structure is at least partially represented as systematic variations in Semantic Space. Thus, we also explored this new line of research.

Overview

This Thesis is divided in three main parts:

- Part II explores the characterization of the dynamic structure of facial expressions for motion stylization and abstraction. To accomplish this, Chapter 2 elucidates the higher-level semantic structure underlying the facial expression space while validating the experimental methodology used throughout the thesis. Then, Chapter 3 extends this research quantitatively and, applies the proposed methodology to MoCap data, in order to provide the necessary tools for the characterization and stylizing of low-level motion information.

- Part III gives a first practical example on how the found SSps can be used together with the gathered Data Bases (DBs) in order to generate new facial expressions. As proof of concept, we used leave-one-out analysis to recover a given expression from the MoCap DB. More specifically, we propose to use the positions of the SSp as weights to interpolate such expression as a weighted combination of the rest of expressions in the DB.

- Finally, Part IV analyzes the mapping between actions and Socio-Emotional space towards giving ECAs an identity expressed not only through expressions but through personality. As a first step, Chapter 5 proves the validity for standardized, validated personality questionnaires to be used to evaluate ECAs psychologically and examines the contribution of each unimodal communication channel indicating that facial expressions are a significant part for personality perception. Finally, Chapter 6 extends the analysis done in Part II widening the number of expressions and actors considered, and analyses the implicit indication of the mapping between the Semantic Space for facial expressions and the personality space.

For all the works presented in this thesis, I am the leading but not the sole author, as all of them where done in cooperation with different colleagues. Thus, at the beginning of each chapter, the contributions of each of the authors is stated when needed after giving a short description of the work to put it in context.

1.3 Contributions and Measurable Results

1.3.1 Publications

The core of this thesis has already been published, mostly in the JCR indexed journal Computer Animation and Virtual Worlds and, together with the publications this thesis originated, I was also able to significantly contribute in some other papers, having in total **three** JCR-indexed journal publications [1, 2, 3] in Computer Animation and Virtual Worlds and ACM Transactions on Applied Perception (TAP) and **six** peer-reviewed conference publications [4, 5, 6, 7, 8, 9].

Besides these publications and the corresponding presentations in conferences for the four in which I was the leading author, I was invited to give five talks:

Image Retargeting. Talk given in Hi-Graphics on the 17th of March of 2013 in Hirschegg-Kleinwalsertal.

The Temporal and Semantic Structure of Dynamic Conversational Facial Expressions. Talk given in Hi-Graphics on the 15th of March of 2014 in Hirschegg-Kleinwalsertal.

Exploring and Travelling the Emospace with MoCap Data. Talk given in Hi-Graphics on the 16th of March of 2015 in Hirschegg-Kleinwalsertal.

Connecting the Semantics of Faces (and of Personality) to the Spatiotemporal Structures of Face Motion. Talk given in the Workshop on Detection of Pain in Facial Expressions on the 19th of February of 2016 in University of Bamberg.

Digital Personality and the Emotional Onion. Invited talk on the 19th of October of 2018 for the Computer Graphics department in TU Braunschweig.

1.3.2 Research Projects and Fellowships

While completing my PhD, I was given the opportunity to participate in two research projects:

DFG Grant (CU 150/2-1) under the title "The Temporal and Semantic Structure of Dynamic Conversational Facial Expressions". This project was led by my advisor, Prof. Dr. habil Douglas W. Cunningham and performed with the cooperation of Prof. Dr. rer. nat. Christian Wallraven (Korea University, Seoul). This fellowship, together with the financial support of the Brandenburg Technical University (BTU) allowed the realization of the work here presented.

"Interaction between new technologies for combined freight transport with the design and operation of railway facilities" financed by the Karl-Vossloh-Stiftung. I was a researcher and designer for this project as a member of the Computer Graphics Team from the BTU Cottbus-Senftenberg.

1.3.3 Professional Service

I have been able to serve the research community as a reviewer for five journals (Image and Vision Computing, IEEE Transactions on Image Processing, ACM Transactions on Applied Perception, Signal Image and Video Processing and IEEE Computer Graphics and Applications) and two international conferences (Eurographics and the ACM Symposium on Applied Perception (SAP)). Several international conferences bestowed me the privilege to contribute; I was part of the local organizing committee for the Eurographics

Symposium on Rendering (EGSR) 2013, served as graphic designer for Informatik 2015 and ACM SAP 2017. For the latter, I also had the honor to be the Posters Chair.

1.3.4 Academical Experience

During my PhD student's years, I also got the chance to gather some teaching experience, both in supervised and autonomous ways. I was a teaching assistant for three different lectures, where I co-guided the students for the seminars "Perception and Sensation for Computer Scientist" [2012–2016] and "Models of Human Perception" [2015–2016] and took care of preparing and supervising the exercises for "Designing and Understanding Psychological Experiments" [2014–2018]. Once I got the formation in conducting the two seminars together with my advisor, I was given the opportunity to teach both of them on my own following the program given in the previous years [2016-2017]. This academic training allowed me to become lecturer for Theoretical Computer Science (Theoretische Informatik), where I assumed all teaching responsibilities, from creating and conducting the program to preparing and correcting the exams [2014–2016].

Supervised Theses

Between 2013 and 2018, I have had the honor of co-advising more than twenty BSc and MSc theses, covering diverse topics from Computer Graphics:

> Ilka Klug (2014): *Multiresolution Mesh Morphing* (Bachelorarbeit im Studiengang Informations- und Medientechnik).
>
> Sophie Baschinski (2014): *Semiautomatische Vernetzung eines Gesichtsmodells mit einem beliebigen Modell eines menschlichen Körpers* (Bachelorarbeit im Studiengang Informations- und Medientechnik).
>
> Thomas Schulze (2014): *Überblick und Bewertung von Prosodie in 'Text-to-Speech'–Systemen* (Bachelorarbeit im Studiengang Informations- und Medientechnik).
>
> Christian Borck (2014): *Automatisches Deblurring von Fotografien unter Verwendung einer Patch-basierten Analyse von Lichtfeldern* (Bachelorarbeit im Studiengang Informations- und Medientechnik).

Oliver Költzsch (2014): *Framework für eine allgemeine multi-level Analyse von Embodied Conversational Agents* (Bachelorarbeit im Studiengang Wirtschaftsingenieurwesen Informatik).

Pascal Glang (2014): *Auswirkungen Nicht-Fotorealistischer Rendering-Stile auf die wahrgenommenen Charaktereigenschaften eines 3D-Gesichtsmodells* (Masterarbeit im Studiengang Informations- und Medientechnik).

Katharina Legde (2014): *Entwicklung eines Affective Talking Head* (Masterarbeit im Studiengang Informations- und Medientechnik).

Martin Schorradt (2014): *Integration und Evaluation verschiedener Emotionen in einem artikulatorischen Sprachsynthesesystem* (Bachelorarbeit im Studiengang Informatik).

Wei Lu (2014): *Performance-Driven Facial Animation using Motion Capture Data and a 3D Head Model* (Bachelorarbeit im Studiengang Informations- und Medientechnik).

David März (2015): *Implementierung einer dynamischen Amöbe für Segmentierung in einem Image Inpainting Algorithmus* (Bachelorarbeit im Studiengang Informations- und Medientechnik).

Thomas Kantor (2015): *Flexible Bewegungssynthese zur Animation des Körpers eines Virtuellen Agenten* (Masterarbeit im Studiengang Informations- und Medientechnik).

Wei Lu (2016): *Eine analysebasierte Synthese zur Gewinnung von menschlichen 3D Bewegung aus echten und animierten 2D Video Sequenzen* (Masterarbeit im Studiengang Informations- und Medientechnik).

Philipp Hahn (2016): *A Blender based Affective Talking Head with a variable degree of visual realism* (Masterarbeit im Studiengang Informatik).

Artjom Sosin (2017): *Implementierung parametrischer Bewegungsgraphen auf Basis von Motion-Capture-Daten zur Animation virtueller Agenten.* (Bachelorarbeit im Studiengang Informations- und Medientechnik).

Maximilian Mühle (2017): *Erweiterung eines Image Inpainting Algorithmus mit Hilfe einer dynamischen Amöbe* (Bachelorarbeit im Studiengang Informatik).

Marco Menzel (2017): *Rekonstruktion von 3D Blickrichtungen aus 2D Videos (Reconstruction of 3D Eye-Gaze Direction from 2D Videos)* (Bachelorarbeit im Studiengang Informatik).

Martin Karras (2017): *Rekonstruktion von 3D Kopf- und Gesichtsbewegung aus einem 2D Video (Recovering 3D Head- and Facemotion from a 2D Video)* (Masterarbeit im Studiengang Informations- und Medientechnik).

Baoqiang Yang (2017): *Mesh Reconstruction using Differential Geometry* (Bachelorarbeit im Studiengang Informations- und Medientechnik).

Tobias Wacker (2017): *Rekonstruktion von Motion Capture Daten* (Bachelorarbeit im Studiengang Informatik).

Martin Schorradt (2018): *Development of a Method for Lossless Prosody Isolation* (Masterarbeit im Studiengang Informatik).

Alexej Stumpf (2018): *Leichte und robuste 3D-Objekte* (Bachelorarbeit im Studiengang Informatik).

Part II

Facial Expressions

In this part we introduce the use of Semantic Spaces to characterize facial expressions. We recover the Semantic Space for conversational facial expressions using the following as stimuli: emotional words, video recordings of people while showing expressions, and the videos of point clouds for the corresponding Motion Capture synchronized recordings. The high correlation between these spaces, allows us to make the abstraction from an idea of an expression given by a word – like "Pleasantly Surprised" – to specific facial movements determined by Motion Capture data. This, together with their continuous nature, makes Semantic Spaces a promising tool for semantic-driven facial animation.

THE SEMANTIC SPACE FOR FACIAL COMMUNICATION

2

We can learn a lot about someone by watching their facial expressions and body language. Harnessing these aspects of non-verbal communication can lend artificial communication agents greater depth and realism, but requires a sound understanding of the relationship between cognition and expressive behaviour. In this chapter, we extend a traditionally word-based methodology to use actual videos and then extract the semantic/cognitive space of facial expressions. We find that depending on the specific set of expressions used, either a four- or a two-dimensional space is needed to describe the variance in the stimuli. The shape and structure of the 4D and 2D spaces are related to each other and very stable across methodological changes. The results show that there is considerable variance between how different people express the same emotion. The recovered space can well capture the full range of facial communication and is very suitable for semantic-driven facial animation.

An edited version of this work is published in a special issue of the *Computer Animation and Virtual Worlds Journal by Wiley* which was presented in the 27*th International Conference on Computer Animation and Social Agents (CASA 2014)*, hold on May $26 - 28$, 2014 at the University of Houston, Houston, Texas, USA. The co-authors of the mentioned paper were Prof. Dr. Christian Wallraven (Department of Brain and Cognitive Engineering, Korea University), and the advisor of this thesis, Prof. Dr. Douglas W. Cunningham (Graphic Systems Department, BTU Cottbus-Senftenberg), they both gave advice concerning ideas and content of the paper.

S. Castillo, C. Wallraven and D. W. Cunningham
The Semantic Space for Facial Communication
Computer Animation & Virtual Worlds, 25: 223-231. CASA 2014

2.1 Introduction

We are all experts to one degree or another in many different forms of natural language communication. Face-to-face conversation is generally the one we are best at, and is generally one of the most powerful. As has been pointed out many times, any computer system that can handle the subtleties of natural conversations will be instantly usable by almost anyone. Of course, this is not an easy task and is the subject of a host of research in many different scientific fields.

One of the hallmarks of natural conversations is also one of the things that makes it so complex: it uses many different physical channels simultaneously (e.g., the voice, body, and face/head). Moreover, any given physical channel is used to simultaneously accomplish different tasks. For example, we can use our voice to utter several words, to emphasize a certain word in that utterance, and to indicate that the utterance is a question – all at the same time. To capture this complexity, de Ruiter et al [22] introduced the concept of the semiotic channel. A semiotic channel is a set of behaviours whose elements (1) cannot be performed simultaneously with each other and (2) can be performed simultaneously with (almost) all behavioural elements in other semiotic channels.

It has since been shown that "non-verbal" semiotic channels such as can be found in the physical channels of hand gestures, body language, eye gaze, and facial expressions can serve many different functions, including conveying a concept (either alone or in concert with another channel; see, e.g., the work of Paul Ekman [14]), modify the information in another channel (e.g., the research of Condon and Ogston [16]), provide emphasis (e.g., Krahmer et al.'s work [15]), and control conversational flow (e.g., V. H. Yngve's research [17]). Furthermore, considerable evidence exists that in conflict situations, the information in these "non-verbal" channels tend to be given more weight [12, 23, 24, 25]. In a particularly interesting example, Archer and Akert [23] systematically manipulated the emotional content of different semiotic channels and showed that when the verbal (or semantic) content and the "non-verbal" conflict, most people place considerably more weight on the non-verbal signals.

2.1.1 Semantic Modelling for Facial Animation

The various roles of "non-verbal" semiotic channels take on additional weight for behavioural animation when one considers techniques like motion style transfer [26]. In motion style transfer, the "style" (such as sneaky, happy, or bold) of a point light walker's motion is copied to another point light walker without altering the content of the second walker's motion. It should be possible, by analogy, to tweak the behaviour of an agent to alter the tone of a message without altering the semantics the message. That is, we should be able to give the agent a visible personality.

One of the first to concretely formulate an idea along these lines in the field of computer graphics were Funge, Tu, and Terzopoulos [27], who discussed cognitive modelling for behavioural animation. While many people have developed a number of impressive forms of cognitive-based behavioural modelling (for a recent overview, see the work of Kapadia et al. [28]), one that stands out is Badler and colleague's work combining the theories from personality psychology, dance, and character animation [28, 29]. Two representational systems are at the core of their work. One is a system called EMOTE [30], which is derived from Laban Movement Analysis (which describes body movements). The second system is the personality model OCEAN [31], which uses five dimensions (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism) to represent human personality. By defining a mapping between these two representational systems, high-level personality traits can be used to alter, choose, or direct the aspects of a character's movement.

In order to extend such cognitive-based behavioural animation to affective communication agents, the *facial equivalents* of Laban Motion Analysis and OCEAN are needed, along with a mapping between them. There are a number of successful systems for describing facial expressions, although most of them focus not on motion but on static deformations (for a review of facial coding systems, see the State-of-the-Art written by Vinay Bettadapura [32]). There are also a few cognitive models of emotions. These models, however, focus on *emotional states* rather than on *communicative intention*. Several researchers (e.g., [33, 34]) have explored a few mappings between existing models of emotions and agents, but it is unclear how well models of emotional state are able to explain the complexity of face-to-face communication in general – the most trivial example being non-emotional communicative expressions, such as agreement and thinking. Since our goal is to go beyond facial expressions of emotions and describe the full

space of communicative facial gestures [10, 35], we therefore address in this chapter the problem of defining and measuring the full Semantic Space of facial communication, hence laying the groundwork for later cognitive-based facial animation of communication agents.

2.1.2 Emotion Space

Within the emotion literature, there are two core categories of models for describing the space of emotions: two-dimensional models and four-dimensional models [36]. These models are based on theoretical considerations as well as empirical analysis, most of which is based on Osgood's research [20]. Osgood essentially wanted to understand how people could compare apples and oranges. To do this, he developed the semantic differential method. Specifically, he asked people to rate many words along many Likert scales, with each end of a scale being anchored by a pair of opposites (e.g., good versus bad, strong versus weak). The covariation of the ratings was then examined using either PCA or factor analysis to extract the underlying structure (for more on Likert scales, semantic differentials, and factor analysis, we refer the reader once again to the work of Cunningham and Wallraven [21]). Osgood found that regardless of what concepts people were rating, the same three dimensions showed up and together they were sufficient to capture over 70% of the variance in the ratings. These three dimensions are Evaluation, Potency, and Activity (EPA). In some situations, the fourth dimension (Predictability) is also important.

Many models of emotion extract all four dimensions (see, e.g., [18]). Some, however, obtain only two dimensions, which are usually called Valence and Arousal (see, e.g., [37, 38, 39]). Notice that the first and third dimensions of the 4D EPA-based solution are similar to Valence and Arousal.

In a recent, very thorough, empirical examination of emotional words, Fontaine and colleagues had a number of participants from three cultures imagine that a person was experiencing one of 24 core emotions. The participants then had to rate how likely 144 other behaviours/emotions (such as breathing fast or wanting to be close to someone else) were. They found that 75% of the variance along these 144 scales can be explained with four dimensions, which strongly resemble the four EPAP dimensions [18]. Note, that these experiments used imagined situations rather than actual photos or videos of

emotions as stimuli. In everyday face-to-face communication, however, we constantly try to infer someone's communicative and emotional state from visual and acoustic signals such as facial gestures. Hence, here we are interested in performing a semantic differential analysis using visual stimuli (facial expressions) instead of imagined situations. For this, we first replicate Fontaine et al.'s results using a slightly modified methodology, and then examine the Semantic Space of facial communication that is obtained when actual videos are used instead of words.

2.2 General Methods

In the following we describe the scales and general psychophysical methodology. Since all participants were German native speakers, all stimuli and instructions were in this language.

2.2.1 Scales

We based our scales on Fontaine et al.'s [18] 144 unipolar scales. Since three scales per dimension are sufficient [20], we selected 12 scales that were highly correlated with the four emotional dimensions. It was critical that none of the scales specified a visible behavior (such as "breathed heavily"), since these would be obvious when videos are used as stimuli. Since the majority of work in semantic differentials is with bipolar scales (anchored at both ends by adjectives), we converted the unipolar scales to bipolar scales by adding the opposite to the other side of the scale. In many cases, this opposite was also one of Fontaine's 144 scales. If so, then we required that opposite to correlate highly with the same factor. The 12 scales and their correlation to the factors as derived in the work of Fontaine et al. [18] can be found in Table 2.1.

2.2.2 Procedure and Design

After filling out an informed consent form, participants were placed one at a time in a semi-dark room roughly 0.5 m in front of a 24" LED monitor (at a resolution of $1920x1080$). The experiment was controlled by Psychophysics Toolbox Version 3.0.11

FACTOR	SCALE	
	ID	**Anchors**
F1: Evaluation	1	Felt Positive -Felt negative (Fühlte sich positiv - Fühlte sich negativ)
	5	Felt liberated or freed - Felt inhibited or blocked (Fühlte sich befreit - Fühlte sich gehemmt oder blockiert)
	9	Wanted to be near or close to people or things - Wanted to keep or push things away (Wollte nah an den Leuten oder Dingen sein - Wollte Abstand halten)
F2: Potency	2	Felt strong - Felt weak (Fühlte sich stark - Fühlte sich schwach)
	6	Felt dominant - Felt submissive (Fühlte sich dominant - Fühlte sich unterlegen)
	10	Wanted to tackle the situation - Lacked the motivation to do anything (Wollte die Situation anpacken - Motivationslos)
F3: Activity	3	Felt restless - Felt calm (Fühlte sich rastlos oder unruhig - Fühlte sich ruhig)
	7	Heartbeat got faster - Heartbeat slowed down (Schnellerer Herzschlag - Langsamerer Herzschlag)
	11	Breathing got faster - Breathing slowed down (Schnelleres Atmen - Langsameres Atmen)
F4: Predictability	4	Caused by an unpredictable event - Caused by a predictable event (Durch ein unvorhersehbares Ereignis verursacht - Durch ein vorhersehbares Ereignis verursacht)
	8	Experienced the emotional state for a short time - Experienced the emotional state for a long time (Emotion dauerte kurz an - Emotion dauerte lang an)
	12	Caused by chance - Predictable cause (Geschah zufällig - Geschah vorhersehbar)

Table 2.1: The 12 scales considered in this study grouped by dimension. The German equivalent that was used in the experiment is displayed in parentheses.

(PTB-3) [40, 41, 42]. The participants then were presented with a screen describing in detail the instructions, and were given another chance to ask questions.

At the start of each trial, participants were presented with a word (Experiment 1) or a video (Experiment 2) on the right of the monitor and a scale from 1 to 7 on the left (see Figure 2.1). The ends of this scale were represented with a pair of opposing words. At the top of the screen the main question was always displayed: How likely is it that these emotional features also occurred? ("Wie wahrscheinlich ist es, dass diese Merkmale auch vorkamen?"). Participants were explicitly instructed to not simply rate the nature of the expression in terms of the scale (for example, a person who is strongly depressed can feel very weak). The participants were able to enter their answers by clicking on the desired level of the scale using the mouse, once a response was entered, the next trial started. Each stimulus was rated on all 12 scales before a new stimulus was shown. The order of the stimuli was random, with each participant receiving a different random order. Each participant reported normal or corrected-to-normal vision, and was compensated at 8€ per hour.

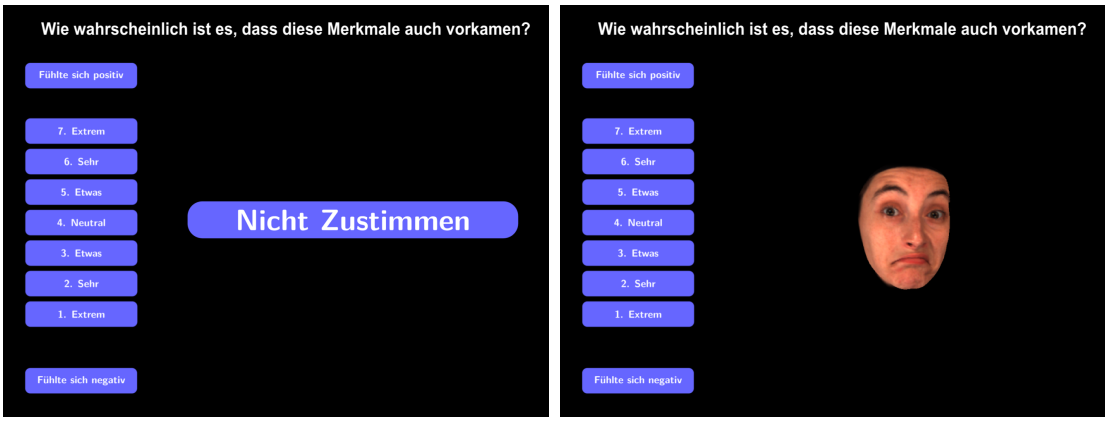

Figure 2.1: Snapshot of one trial from Experiment 1 (left) and Experiment 2 (right). Note that the same interface was used (with different stimuli) in both experiments.

2.3 Experiment 1: Emotional Words

Here we validate that our modifications to Fontaine et al.'s [18] design did not affect the results. Furthermore, we add six new words, to match the additional, conversational facial expressions that will be used in the second experiment.

2.3.1 Methods

A total of 11 people participated (age range $20 - 32$; 6 females). The mean time to complete the experiment was 37 minutes. The stimuli for this experiment were 30 words (see Table 2.2) describing a feeling or mental state. Twenty-four of these are the same used by Fontaine and colleagues [18] and represent prototypical emotion terms widely used in daily life and emotional research. We expanded this set with six additional words describing additional conversational reactions.

Prototypical		
Anger (Wut)	Anxiety (Beklemmung)	Being Hurt (Verletzt werden)
Compassion (Mitleid)	Contempt (Verachtung)	Contentment (Zufriedenheit)
Despair (Hoffnungslosigkeit)	Disappointment (Enttäuschung)	Disgust (Ekel)
Fear (Angst)	Guilt (Schuld)	Happy (Glücklich)
Hate (Hass)	Interest (Interesse)	Irritation (Genervtheit)
Jealousy (Neid)	Joy (Freude)	Love (Liebe)
Pleasure (Vergnügen)	Pride (Stolz)	Sadness (Traurigkeit)
Shame (Scham)	Stress (Stress)	Surprise (Überraschung)
Conversational		
Agree (Zustimmen)	Clueless (Unwissend)	Confused (Verwirrt)
Disagree (Nicht Zustimmen)	Pleasant Surprise (Angenehm Überrascht)	Thinking (Nachdenklich)

Table 2.2: The 30 words considered in this study. Their German equivalents are shown in parentheses.

2.3.2 Results and Discussions

There are a number of methods for determining how many factors are sufficient to explain the variance in the data, each with its advantages and disadvantages (see [43, 44]). In general, it is advised to look at a number of criteria. For the results of Experiment 1, the Kaiser criteria, parallel analysis, and the optimal coordinates all suggest that three factors are needed. The Chi-Squared tests, the explained variance, and theoretical reasons suggest a four-factor solution is needed. So, we will look at both the three and four factor solutions.

Overall, both the three- and the four-dimensional solutions are highly reminiscent of EPA space. In the four-dimensional (4D) solution we get the factor loadings shown in Table 2.3, with Factors 1, 2, 3, and 4 explaining 32.7%, 17.4%, 17.4%, and 16.4% of the variance, respectively. The factor analysis successfully recovers the desired EPA dimensions from the scales (all cells in Table 2.3 which are both gray-shaded and bold-font), with the exception of scale 3. It turns out that the negative side of scale 3 ("Felt calm") belonged to the Evaluative dimension and loads onto it here as well. The different order of the four factors between the present experiment and that of Fontaine et al. – along with the difference in the amount of variance that they explain – can be easily accounted for by the fact that all of our 12 scales correlated well with the chosen dimensions, whereas many of Fontaine et al.'s 144 scales correlated weakly at best with any factor. The three dimensional solution is the same as the 4D solution, except that Evaluation and Potency are fused.

Scale ID	**F1**	**F2**	**F3**	**F4**
1	**1.037**	0.069	-0.023	-0.086
2	0.287	0.056	-0.005	**0.743**
3	-0.862	0.330	0.006	-0.015
4	-0.105	-0.031	**0.908**	-0.048
5	**0.820**	0.006	-0.025	0.217
6	-0.050	-0.111	0.008	**1.040**
7	0.021	**0.981**	0.018	-0.042
8	0.197	0.072	**0.528**	0.103
9	**1.061**	0.072	0.002	-0.180
10	0.421	0.173	-0.086	**0.475**
11	-0.026	**0.980**	-0.031	-0.028
12	-0.004	0.005	**0.989**	0.012

Table 2.3: Factor loadings. The gray-shaded cells show the significant contributions found by Fontaine et al. [18] while the bold values show the ones recovered from our solution.

We can now compare the location of each emotional word in the 4D space (see Figures 2.2, 2.3, and 2.4) to Fontaine et al.'s [18] results using a procrustes analysis. The standard distance measure is the Sum of Squared Errors (SSq) between the two matrices, which in this case yields a distance $d = .437$. Since the correlation of the two matrices can be calculated from the SSq ($r^2 = 1 - SSq$) [45], we see that they are correlated at $r = .75$, which is rather high. A part of the deviation between the two sets of results might be due to intercultural variance [18]. Some of the deviation might also be due to our use of fewer scales.

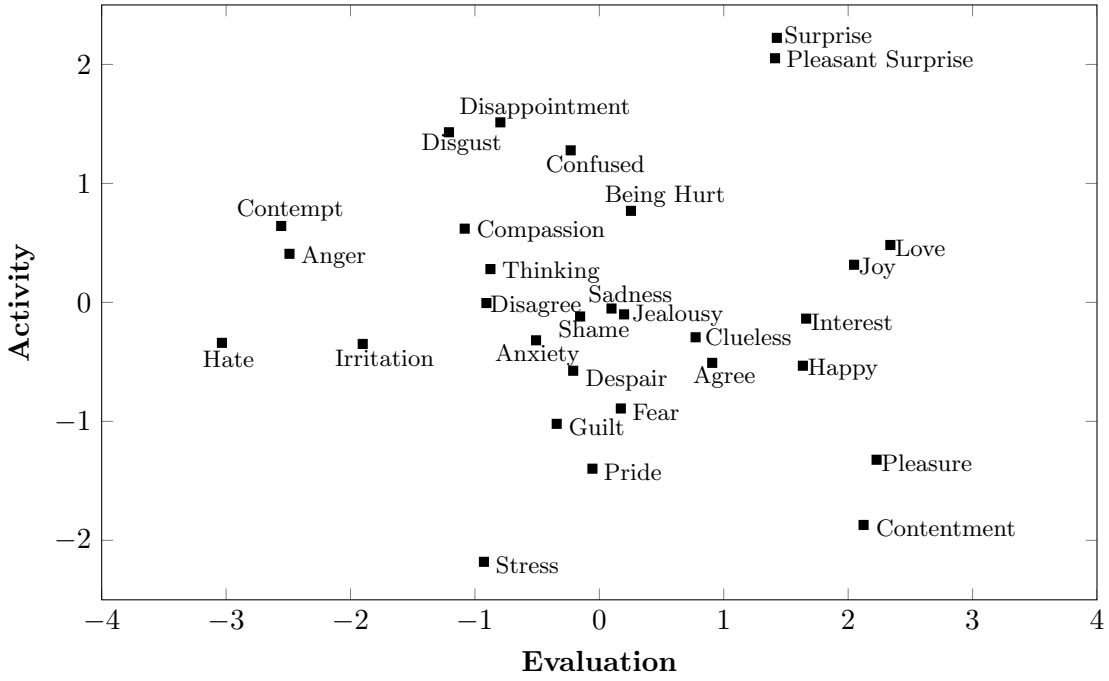

Figure 2.2: Coordinates along Evaluation and Activity for the 30 words.

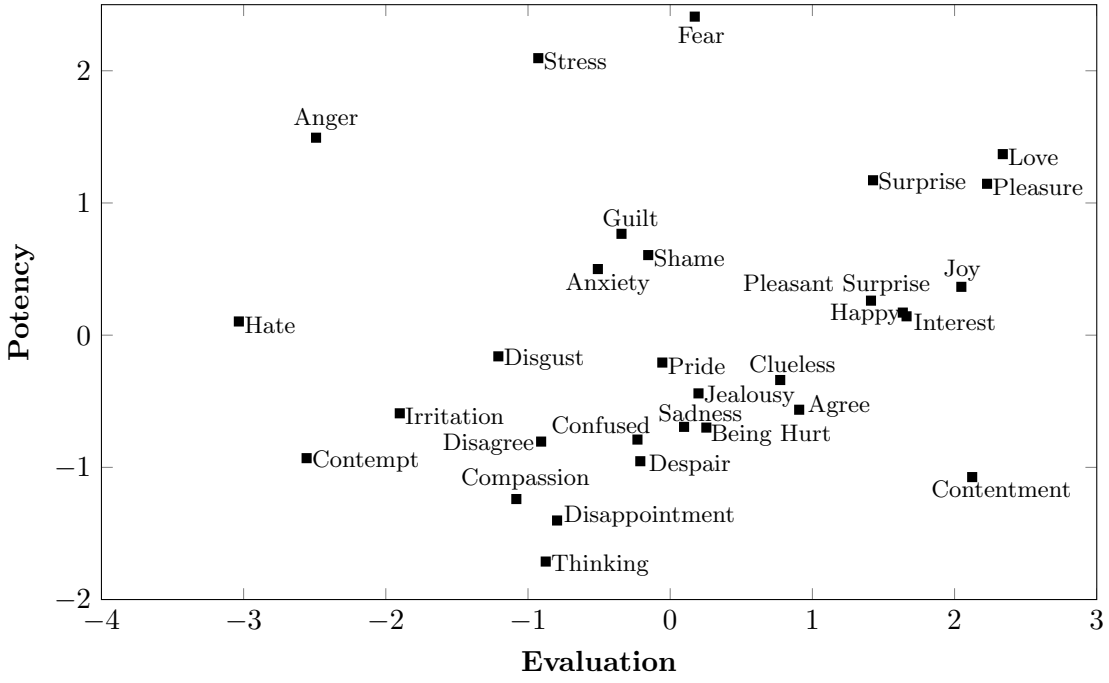

Figure 2.3: Coordinates along Evaluation and Potency for the 30 words.

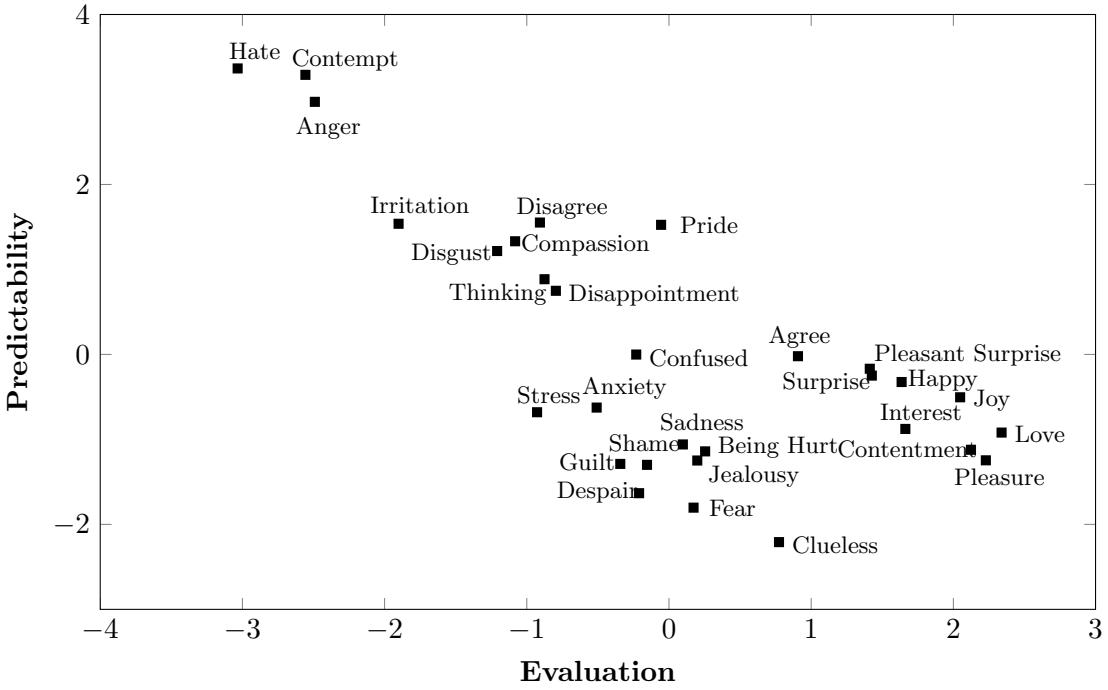

Figure 2.4: Coordinates along Evaluation and Predictability for the 30 words.

In anticipation of Experiment 2, we re-ran the analyses using only the nine words representing the nine expressions in Experiment 2. Since a factor analysis requires more stimuli than scales, we used a PCA instead. All criterion suggest a 2D solution, which fuses Evaluation and Potency as well as Activity and Predictability. Note that this resembles the classic Valence-Arousal model of emotion. The difference between the results with all 30 words and the results with just these nine also serves to reinforce the fact that factor analysis and PCA can only detect variance if it is present in the stimuli.

2.4 Experiment 2: Video as Stimuli.

In Experiment 1, participants imagined that someone was experiencing a given emotion. One of the core disadvantages of imagination is that there are many sources of potential variance, all of which serve to mask the true nature of the underlying Semantic Space. For example, imagination is based on personal experience, and the exposure to a given emotion will differ between participants. More critically, the emotional words used in

Experiment 1 are almost all basic-level categories (i.e., an average level of specificity, the most common level elicited in everyday questions; [46]). The imagined emotion, however, will usually be a specific variant (subordinate level category): for example, one participant might choose to imagine "elation" for happy, while another might choose "satiated". Even if two people choose the same subordinate category, the intensity imagined might be very different.

In order to obtain a clearer picture of the Semantic Space of emotional expressions, it will be useful to more carefully control what is being rated. Moreover, since the final goal of this thesis is to construct a Semantic Space whose elements will eventually be mapped to specific facial deformations or facial motions, examining the Semantic Space of actual facial expressions is the logical next step. Thus, in this experiment, we replicate the previous experiment using videos as stimuli.

2.4.1 Methods

Ten people participated in the experiment (age range $20-28$; 5 females). The stimuli were from the small MPI facial expression database [10], they were recorded using a method acting protocol for six people (two male, four female) (see Figure 2.5) and consisted of nine expressions (agree, disagree, happy, sad, clueless, thinking, confusion, disgust and pleasantly surprised; see Figure 2.6). The mean time to complete the experiment was 73 minutes.

Figure 2.5: The six individuals recorded in the small MPI facial expression database [10] used in this study.

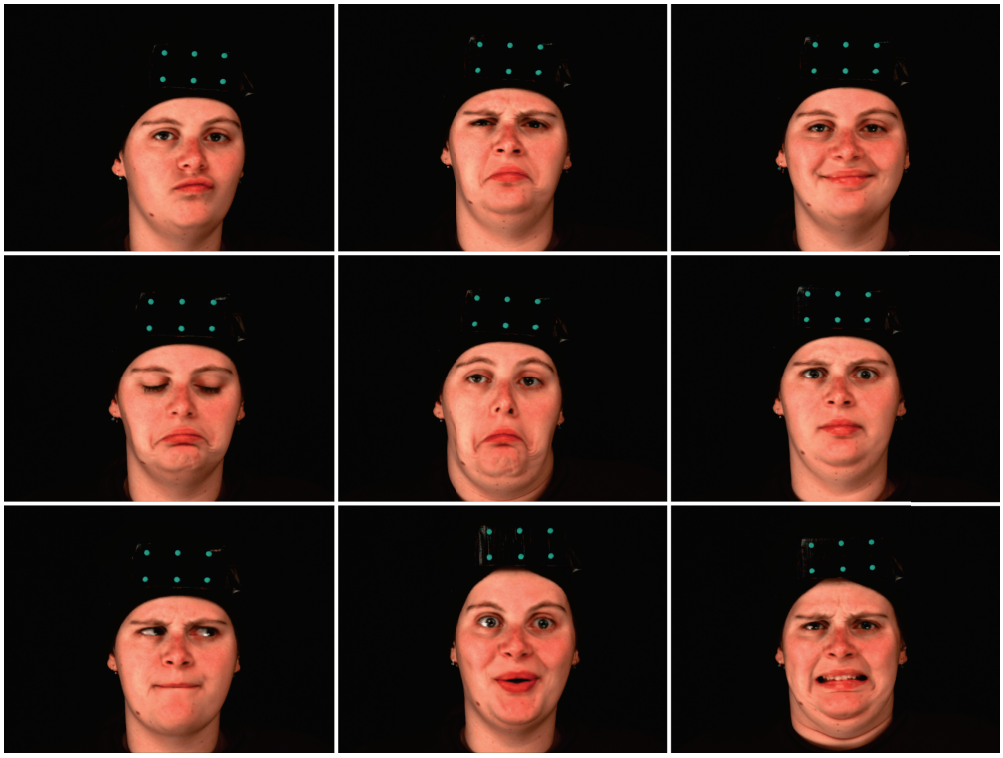

Figure 2.6: Static peak frames of the nine expressions for one person (the markers above the head are used for tracking).

2.4.2 Results and Discussions

The Kaiser criterion, parallel analysis, optimal coordinates, and acceleration factor all argue for a two factor solution. Explained variance argues for a two or a three factor solution, for 80% and 90%, respectively. Chi Squared argues for a three factor solution.

The two factor solution is the same as found in Experiment 1 for the nine words representing the expressions used here. Specifically, Evaluation and Potency are fused into a general Valence dimension, while Activity and Predictability are fused into a general Arousal dimension. The three dimensional solution shows essentially the same loadings, with very minor differences.

The location of the different emotions, along with the variance due to actors, can be found in Figure 2.7 where the ellipse around each square is the Standard Error of the Mean (SEM) for the distance of each actor to the mean expression. As can be seen, there is considerable variance between the different actors for each expression. This means that

there is considerable flexibility in *how* one expresses a given emotion or state, and some of this flexibility surely related to the personality of the individual actors (for example, same actors tended to be more on the negative side of all expressions, etc.). This is probably better illustrated in Figure 2.8, where the average among all expression for every given actor is visualized together with the averaged center of all emotions among all actors (which, logically, yields the origin of coordinates). This image shows the shift an actor gives with his or her interpretations of the generic set of emotions (i.e., the average location for a given emotion). There is also considerable variance between the different emotions, with some obvious clustering. This implies that, it should be possible to create a facial motion that may not be directly identifiable, but will give the correct tone or personality impression.

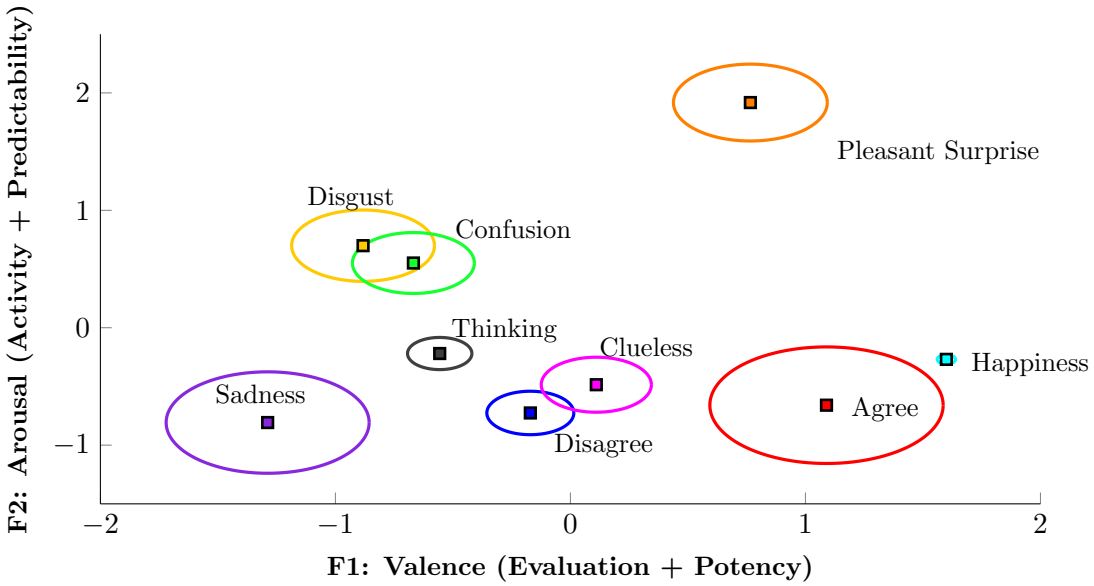

Figure 2.7: Coordinates for the facial expressions. The squares represent the video results and the radius of the ellipse around each emotion is the SEM for Euclidean distance of each actor to the mean expression.

To more directly compare the results of Experiment 2 to those in Experiment 1, we averaged across all actors for each expression and then ran a PCA. All criteria suggest a 2D solution, with the same loadings as for the complete analysis. A procrustes analysis gives a distance between the two matrices of $d = .25$, yielding a correlation of $r = .87$ between them.

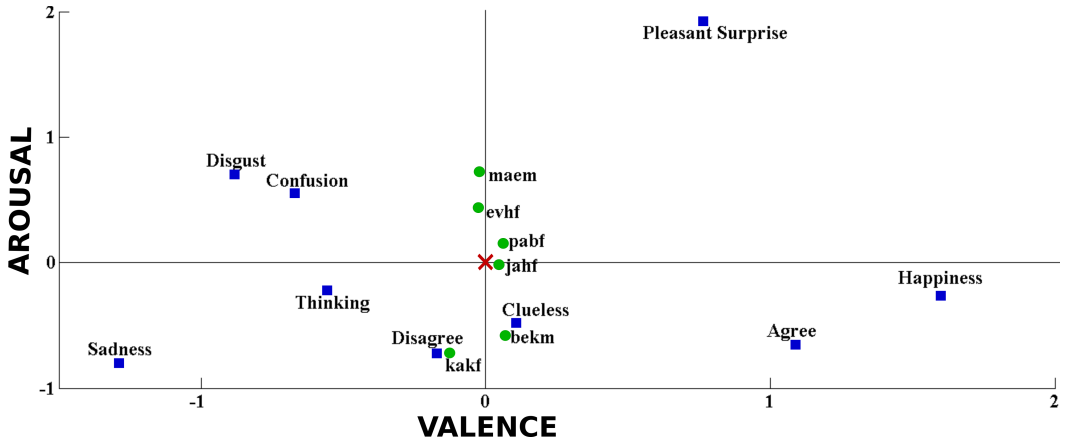

Figure 2.8: Coordinates for the facial expressions and their averaged center (red cross). The green spots represent the mean position for all emotions among a given actor, illustrating his/her distance to the expressions' center.

A brief examination of the Euclidean distances for the individual emotions shows that although eight of the nine expressions are indeed close to each other, clueless is on located on opposite sides of the 2D space for Experiments 1 and 2, even after procrustes rotation. It seems that the participants interpreted the word *Clueless (Unwissend)* and the videos of clueless expressions differently. We consequently repeated the analyses without *Clueless*. The PCAs still yield a 2D solution (for both the eight words and the eight videos). A procrustes analysis of this subset yields a distance of $d = .06$, and an impressive correlation of $r = .97$. Moreover, the transformation required to align the two spaces is very minor, consisting solely of a 2 degree rotation and a scaling of .97 (with no reflection). Given the trivial nature of the transformation, we examined the distance between the two matrices *without rotation or scaling* (i.e., the original matrices), and found a distance of $d = .636$, giving a correlation of $r = .968$. This suggests that the relationship between words and videos is very similar (at least for eight of the nine expressions). More critically, it demonstrates that the shape of the recovered space is rather stable, making it suitable Semantic Space for semantic-driven facial animation.

2.5 General Conclusions

For Experiment 1, we ran a standard semantic differential task, using scales and emotion words derived from Fontaine et al. [18], along with six new conversational expression

words. A factor analysis of the results successfully recovered the same four-dimensional (4D) space found by Fontaine et al., corresponding to Evaluation, Potency, Activity, and Predictability. The location of the individual words within the 4D space was similar to previous work as well.

In Experiment 2, we used the same task with video sequences of nine conversational expressions, each recorded from six people. Factor analysis found that two dimensions were sufficient to describe the variance. The two dimensions fuse Evaluation and Potency as well as Activity and Predictability. When Experiment 1 was re-analyzed looking just at the nine words that describe the nine expressions in Experiment 2, the words also yield a two-dimensional solution, with the location of the expressions being akin. In other words, the space for words and expressions seems to be very similar.

Naturally, a mapping between the Semantic Space and facial *motions* must still be found. Ahn and colleagues have done some initial work in this direction using theoretical considerations to relate facial action units (FACs; [14]) to a two-dimensional circumplex representation of emotions dimensions [33]. Of course, given the critical nature of temporal changes in expressions [47] combined with the findings in transfer of motion style, it would seem to be more appropriate to use a spatio-temporal description of facial expressions. Likewise, it might be advisable to empirically derive the mapping.

Interestingly, the use of videos allowed us to have different versions of the same emotion. Since different people have different personalities, they will perform the same expression differently. Moreover, it has been previously shown that no one is good at all expressions [10]. Thus, an examination of how a given person's expressions differ from the average can provide insights into their profile. For example, the actress KAKF is more active than average for all active expressions and more passive for all passive expressions, suggesting a tendency to exaggerate or to use higher intensities. The more extreme the average expression is, the greater the degree by which KAKF exaggerates. Using more additional expressions and a measure of the actor's personality (or perceived personality traits), we should be able to develop a solid mapping between OCEAN and the Semantic Space for facial expressions, completing the pipeline from personality space through semantic expression space to actual behaviour and facial expressions.

THE SEMANTIC SPACE FOR MOTION-CAPTURED FACIAL EXPRESSIONS

3

During our daily lives we convey information verbally and non-verbally. Most of the affective meaning of a message is transferred with the help of facial expressions and, thereby, when trying to establish a realistic human-like virtual character, we should pay close attention to the animation. Motion Capture (MoCap) is one of the most common techniques, but due to the wide range of expressions humans use, the recording time and data needed is vast. To address this problem, we propose the use of Semantic Spaces as they help characterizing and positioning expressions by finding a correlation in between them. Thus, in this chapter, we extend our research by providing the semantics spaces underlying real videos and MoCap-Data for a total of 62 conversational expressions. Our results highly correlate with the ones obtained in the previous chapter, showing that our new expressions were correctly recognized. Moreover, these new results can be used to directly project potential new recordings of these 62 expressions on the found spaces.

An edited version of this work is published in a special issue of the *Computer Animation and Virtual Worlds Journal by Wiley* which was presented in the 31*st International Conference on Computer Animation and Social Agents (CASA 2018)*, hold on May 21 – 23, 2018 in Beijing, China. Both co-authors of the mentioned paper belong to the Graphic Systems Department, BTU Cottbus-Senftenberg. Katharina Legde helped with the cleaning of the database and the advisor of this thesis, Prof. Dr. Douglas W. Cunningham, oversaw/guided the project.

S. Castillo, K. Legde and D. W. Cunningham
The Semantic Space for Motion Captured Facial Expressions
Computer Animation & Virtual Worlds, 29: e1823. CASA 2018

3.1 Introduction

Communication is an essential part of our lives. We communicate verbally with the help of our voice or non-verbally through facial expressions, body posture, and hand gestures. Beyond the transmission of the semantic content of a message, we also continuously provide information among these semiotic channels about socio-emotional content like our state of mind, mood, health, feelings, relationships to other people, etc. [48]. Importantly, when interacting with machines, we still tend to use the *apparent* interpersonal behavior cues coming from them [49, 50]. Thereby, it becomes even more important to explicitly address the socio-emotional signals when designing human-like virtual characters.

Previous work has found that 55% of affective meaning is conveyed non-verbally with the help of facial expressions [11, 25]. Given the extensive experience that all people have with facial communication, it should not be surprising that people are experts at both producing and interpreting facial expressions. As such, correctly animating the face is critical.

Many popular facial animation techniques relay on transferring facial expressions from real people to virtual characters via Motion Capture (MoCap). In order to animate a face, the skills of an expert animator and an actor are needed to create, capture, and transfer their interpretation of one expression to the virtual character. A large number of recordings are needed to create a wide range of highly realistic expressions. The discovery of correlations between specific facial motions and different facial expressions (i.e., how a small set of motions can be combined to create many expressions) would help to reduce the amount of MoCap-Data needed. One way to achieved this to create a continuous representation space for socio-emotional meanings (such a reference systems are often called perceptual or Semantic Spaces [18, 20, 36]) along with mapping of individual recordings to their relevant location in the Semantic Space (for more details, see below). Such a method has the added advantage that it provides a mapping from each Motion Capture recording to a common Semantic Space which can be inverted to allow the creation of novel emotions (for example, by using a weighted combination of the existing MoCap-Data). This chapter examines how well MoCap-Data maps into the Semantic Space known to represent the perception of emotional words and video sequences of facial expressions of those emotions as shown in Chapter 2. Therefore, after recovering the Semantic Space for MoCap-Data, similarities to existing Semantic

Spaces need to be explored and the correlation between real videos and MoCap-Data needs to be established. The main contributions of the chapter are: Extending an existing Semantic Space of real videos presented in the previous chapter to include over 50 new expressions, comparing Semantic Spaces of people with and without an acting background, establishing a Semantic Space for MoCap-Data, and establishing a metric mapping between the perception of facial expressions on the one side and specific words, videos, and MoCap Recordings on the other side.

3.2 General Methods

In the following, we describe the experimental procedures that are common to all experiments of this chapter, and remain consistent with the ones in the previous chapter.

3.2.1 Recordings

A total of 62 expressions (see Table 3.1) were recorded for a total of 10 subjects (5 female) during individual sessions, meaning only one actor at a time. We used the specific expressions, "method acting protocol", and scenarios from the work of Kaulard et al. [35] to ensure the naturalness of the recorded expressions. For examples on the scenarios used to trigger each of the expressions, please see Appendix A. Briefly, the method approach starts with the experimenter describing a real-world situation or scenario to the actor. The actor is asked to imagine that they are in the situation and react normally. On occasion, different scenarios are tried one after the other until the desired socio-emotional message (as seen in the facial motion) is triggered. Once an appropriate scenario is found for a given expression, the actor is asked to again imagine the situation and react normally so that three repetitions of the expression can be recorded. The actors were instructed to return to a neutral expression between repetitions of the reaction. Afterwards, the best recorded repetition for each expression was manually selected. All actors were Spanish, as was the experimenter recording the expressions. Neither the actors nor the experimenter had previous acting experience. Note that this database contains considerably more expressions that the four (neutral, angry, sad, and happy) from a single individual used by Deng et al. [51] to show that naming performance was similar for real videos and MoCap videos.

Facial expressions considered in this study		
Agree `[Agr]`	Agree (Consider) `[AgrCons]`	Agree (Continue) `[AgrCont]`
Agree (Reluctant) `[AgrRel]`	AhaRight `[Aha]`	*Anger* `[Ang]`
Annoyed (Bothering) `[AnnoyBo]`	Annoyed (Rolling Eyes) `[AnnoyRE]`	*Arrogant* `[Arrog]`
Bored `[Bor]`	*Compassion* `[Compa]`	***Confused*** `[Conf]`
Contempt `[Cont]`	***Disagree*** `[Disa]`	Disagree (Reluctant) `[DisaRel]`
Disagree (Considered) `[DisaCons]`	Disbelief `[Disbe]`	***Disgust*** `[Disg]`
Embarrassment `[Emba]`	Evasive `[Evas]`	*Fear* (Oh My God) `[FeOMG]`
Fear (Terror) `[FeTe]`	*Guilt* `[Guilt]`	Impressed `[Impre]`
Insecurity `[Insec]`	***Happy*** (Happy) `[HapLau]`	Happy (Achivement) `[HapAch]`
Happy (Satiated) `[HapSat]`	Happy (SchadenFreude) `[HapSF]`	Imagine (Negative) `[Img-]`
Imagine (Positive) `[Img+]`	Maybe, Not Convinced `[Maybe]`	***Clueless*** (Not Know) `[NotKnow]`
Not Care `[NoCare]`	Not Hear `[NoHear]`	Not Understand `[NoUnd]`
Pain (Felt) `[PainF]`	Pain (Seen) `[PainS]`	Relief `[Reli]`
Sadness `[Sad]`	*Shame* `[Sham]`	Smile (Endearment) `[SmlEnd]`
Smile (Encourage) `[SmlEnc]`	Smile (Flirt) `[SmlFli]`	Smile (Wallace and Gromit) `[SmlWG]`
Smile (Sardonic) `[SmlSar]`	Smile (Sad-Nostalgia) `[SmlSN]`	Smile (Triumphant) `[SmlTri]`
Smile (Uncertain) `[SmlUnc]`	Smile (Reluctant) `[SmlRel]`	Smile (Win) `[SmlWin]`
Smile (Yeah, As If) `[SmlYAI]`	*Surprise* `[Surp=]`	***Pleasant Surprise*** `[Surp+]`
Unpleasant Surprise `[Surp-]`	***Thinking*** (Considering) `[ThCons]`	Thinking (Problem Solving) `[ThPSol]`
Remember (Neutral) `[Remb=]`	Remember (Positive) `[Remb+]`	Remember (Negative) `[Remb-]`
Tired `[Tired]`	Treudoof `[Treud]`	

Table 3.1: The 62 expressions considered in this study. Highlighted expressions are common with the ones used in the experiments in Chapter 2 (italics for words and bold for videos of expressions). The abbreviations in squared brackets are used in the figures of this thesis.

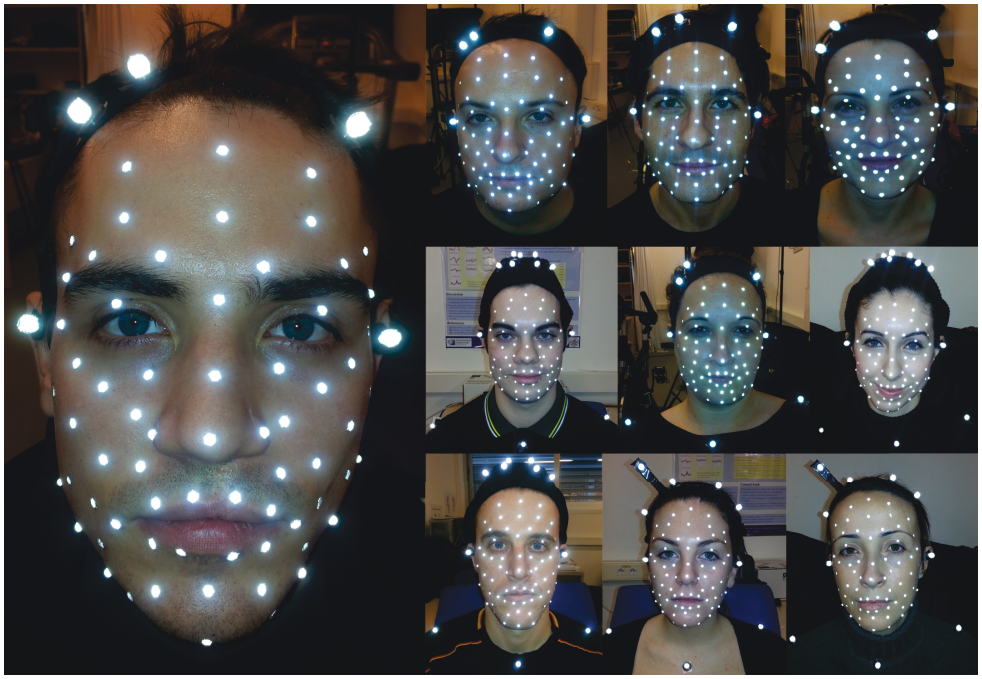

Figure 3.1: Reflecting markers' placement for all the actors (right) with closeup (left).

Apparatus

Before each session, all the actors got 67 reflective markers placed on their face and two in their ears, with the same basic structure that was propose by Breidt et al. [52]. To record and extract the rigid head motion, three different rigs were placed on the actors heads. Half the actors received a hair-band (3 markers for 3 males and 2 females), others received a hat (7 markers for 2 males and 1 female), and the remainder received a diadem (5 markers). Additionally, for 7 of the actors (5 females) we placed 3 extra markers along the collarbone. These marker setups can be seen in Figure 3.1. The average recording time per actor was of 20 minutes for the make-up session, 10 minutes for the calibration of the system, and between 2 and 2.5 hours for the recording.

The general recording setup can be seen in Figure 3.2. The Motion Capture system used in the recordings consisted of 6 VICON MX-F40 Motion Capture cameras with a resolution of 4 Megapixels. Aside from the Motion Capture cameras, the system also had a normal video camera placed on a 30 degrees angle from the front of the actor that synchronously recorded all the expressions. The videos from this camera are the ones used in the experiment described in Section 3.3.

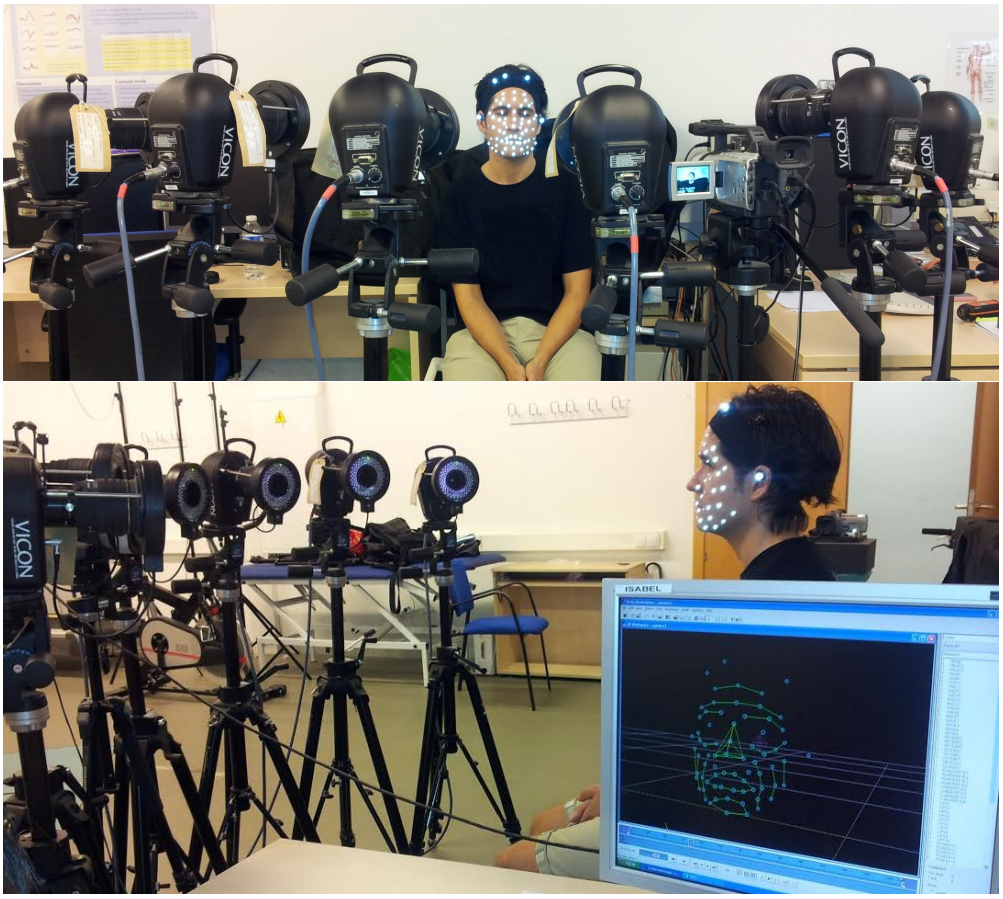

Figure 3.2: The recording setup.

3.2.2 Psychophysical Methodology

We performed two experiments towards recovering the underlying Semantic Spaces on conversational facial expressions. One using real videos as stimuli and the other one using the corresponding MoCap-Data. In the following we describe the scales and general psychophysical methodology used.

Scales

For our experiments, we used the same scales and methodology as proposed in the previous chapter. The twelve scales and the corresponding factors they are designed to correlate with, can be found in Table 3.2.

	SCALE	
FACTOR	**ID**	**Anchors**
Evaluation	1	Felt Positive - Felt negative
	5	Felt liberated or freed - Felt inhibited or blocked
	9	Wanted to be near or close to people or things - Wanted to keep or push things away
Potency	2	Felt strong - Felt weak
	6	Felt dominant - Felt submissive
	10	Wanted to tackle the situation - Lacked the motivation to do anything
Activity	3	Felt restless - Felt calm
	7	Heartbeat got faster - Heartbeat slowed down
	11	Breathing got faster - Breathing slowed down
Predictability	4	Caused by an unpredictable event - Caused by a predictable event
	8	Experienced the emotional state for a short time - Experienced the emotional state for a long time
	12	Caused by chance - Predictable cause

Table 3.2: The twelve scales considered in this study grouped by dimension.

Procedure and Design

Each experiment followed the same general procedure and was controlled by Psychophysics Toolbox Version 3.0.11 (PTB-3) [40, 41, 42]. Since all participants spoke German as their native language, all stimuli and instructions were given in German. All the participants were payed 8€ per hour for their participation. The experiment was described to each participant – but not the research questions behind it – and they were given a chance to ask questions. They then signed an informed consent form. The participants performed the experiment one at a time. They were asked to sit in a semi-dark room roughly 50 cm in front of a 24" LED monitor(at a resolution of $1920x1080$). Each participant was presented with a screen with the instructions for the experiment and, after asking a control question to verify the participant understood the task, the experimenter left the room.

For each trial, the screen showed on the left side a scale from 1 to 7 anchored by one of the pair of terms described in Table 3.2 and, on the right, a video of a real person (Experiment 3; see left image of Figure 3.3) or of a moving point cloud corresponding to the markers with Motion Capture data of one expression (Experiment 4; see right

image of Figure 3.3). The main question for the experiment "How likely is it that these emotional features also occurred?" was always displayed at the top of the screen. The participant needed to select their answer by clicking on one of the displayed numbers before the next trial started. The order of appearance of stimuli was randomized for each participant, but the participant needed to rate one given stimulus in all the 12 scales from Table 3.2 before a new video was selected to be shown.

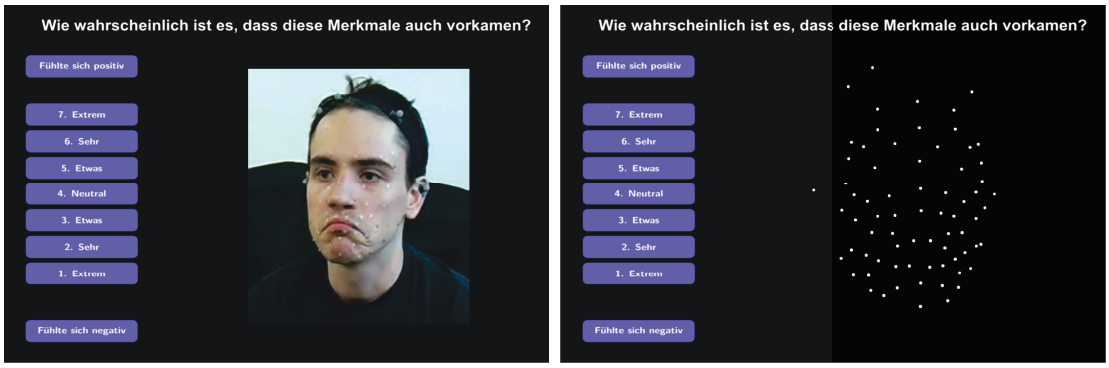

Figure 3.3: Snapshot of one trial from Experiments 3 (left) and 4 (right). Note that the same interface was used (with different stimuli) in both experiments.

3.3 Experiment 3: Real Videos

Consistent with other research on recovering Semantic Spaces [20], we found that the dimensionality of the Semantic Spaces for socio-emotional words depended on the variety and number of stimuli used. The space for 9 words was two dimensional (a classic Valence-Arousal space) while the space for 30 words (including the previous 9) was four dimensional (a standard Evaluation-Power-Activity-Predictability or EPAP space where Evaluation refers to good versus bad, Power refers to strong versus weak, and Activity refers to fast versus slow). The space for 9 real world videos was also 2D, and was identical to the 2D space for words. Here, we replicated Experiment 2 from Section 2.4 presented in Chapter 2 and extended it beyond the previously used 9 conversational expressions to include 53 new expressions (for a total of 62 expressions; see Table 3.1). Note that the goal of this experiment is to expand the stimuli trying to consider as many expressions as possible. Unfortunately, the semantic differential technique is known to become unreliable if there are too many trials (ideally, less than 600 trials should be

used [21]). As the number of trials needed to evaluate 62 expressions on 12 scales is 744, evaluating the videos from more than one actor would surpass the reasonable amount of trials to perform during an experiment. Fortunately, in the previous chapter we already showed that although there is some measurable individual differences for the expressions, the actors were usually tightly clustered. As such, measuring the full space for only one actor should allow us to recovered the full Semantic Space, but his exact locations in that space will only be an approximation of the population average. To maximize the number of expressions in the Semantic Space, we will only use the videos from one representative actor. In particular, we used the data recorded from actor CJCm, whose closeup sample marker setup can be seen on the left of Figure 3.1.

3.3.1 Methods

A total of 17 people participated (age range $22 - 31$; 7 females). The mean time to complete the experiment was 1 hour 55 minutes. The stimuli for this experiment were 62 videos (see Table 3.1) from one actor (CJCm).

3.3.2 Recovering the Semantic Space

The scree test, parallel analysis, optimal coordinates, and acceleration factor criteria agree that two factors are needed while the Kaiser criterion and the explained variance suggest a three-factor solution is needed. We will explore both options (using factor analysis, which is related to the Principal Component Analysis; see [21]).

In the 2D solution (which explains 71.2% of the variance) we get the factor loadings shown in Table 3.3, with Factors 1 and 2 explaining 45.2% and 26% of the variance, respectively (see Figure 3.4). The factor analysis successfully recovers the typical fusion of the EPAP dimensions in the form of Valence (Evaluation and Potency) and Arousal (Activity and Predictability) from the scales (all cells in Table 3.3 which are gray-shaded), with the exception of scale 8 (which goes to Valence instead of Arousal). A possible explanation for this scale to fall into Valence and not into Arousal is that the duration of an expression is sometime linked to intensity and individual differences (and not just to external circumstances) and thus more representative of Evaluation/Potency and not Activity/Predictability.

Scale ID	Factor 1	Factor 2
1	**0.935**	-0.250
2	**0.798**	-0.387
3	-0.340	**0.818**
4	0.181	**0.601**
5	**0.897**	-0.388
6	**0.605**	-0.470
7	0.621	**0.754**
8	**0.296**	-0.107
9	**0.888**	-0.283
10	**0.939**	-0.040
11	0.581	**0.770**
12	0.345	**0.506**

Table 3.3: Factor loadings for the 2D space. The numbers in bold show the significant contributions.

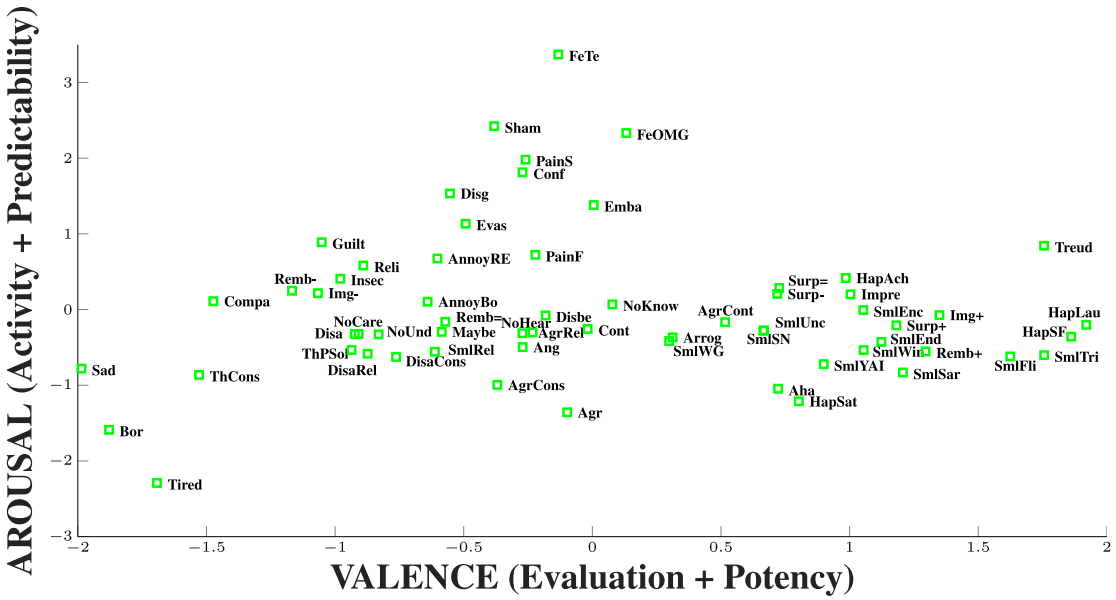

Figure 3.4: Coordinates along Valence (Factor 1) and Arousal (Factor 2) for the 62 real videos.

The three dimensional solution varies a lot depending on the rotation applied. In the case of the direct derivation of the loads, without any rotation, the solution (explaining a total of 76.9% of the variance) is quite similar to the 2D one, with the third dimension consisting solely of scale 3. In this case, the first factor also resembles Valence (scales 1, 2, 5, 6, 8, 9 and 10), explaining 39% of the variance, the second is similar to Arousal (scales 4, 7, 11 and 12), explaining 25.8% and the third factor would stand for 12.1%. This isolation of scale 3 could be due to design issues already mentioned in the previous chapter: while the positive extreme of the scale ('felt strong') belonged and loaded as expected in the Activity dimension, its negative extreme ('felt calm') loaded and belonged to Evaluation.

When applying varimax rotation, the factors resemble the typical EPA space where Predictability is fused with Activity. Specifically, the second factor is almost the same – explains 25.6% of the variance and groups scales 3, 4, 7, 11, and 12. On the other hand, the first and third factors now explain 33.1% and 18.2% of the variance respectively and the corresponding groupings are (1, 5, 9 and 10) and (2, 6 and 8).

3.3.3 Comparison to Previous Experiments

In the previous chapter we showed the Semantic Spaces for words and real videos are identical and the location of the words in that space is essentially identical to the average location (averaged across 6 actors) of videos of those expressions. Experiment 3 (Section 3.3) showed that the the full set of 62 expressions can also be interpreted as a 2D, Valence-Arousal Semantic Space. Here, we examine the relationship between location of expressions in the two experiments. Since Experiment 2 (Section 2.4) only had 9 expressions (from the small MPI facial expression database [10] recorded using a method acting protocol for six actors – two male, four female), we can only check to see if those 9 are the same location.

On the other hand, since we already found a fair amount of individual differences for the specific expressions, we should not expect our single actor to be at exactly the same spot as any of the other actors. As an initial examination as to whether the position of our new actor is similar to the old actors, we projected both the new videos of actor CJCm and the videos from Experiment 2 (Section 2.4) into the same Semantic Space (see in Figure 3.5. Note that since the data are from two different experiments, we

cannot use the factor analysis directly to obtain the scores (i.e., the exact location in Semantic Space). Fortunately, we can recover the space (albeit shifted and scaled by a constant factor) using the loadings of the space since the same scales were used in both experiments. The squares in Figure 3.5 represent the average position of an expression in Experiment 2 (Section 2.4), the triangles are the projections of CJCm videos into the space. The solid ellipses around each expression are the mean of the Euclidean distance of each actor in previous work to the mean position. The dashed ellipses are those means recalculated using all the old actors as well as CJCm.

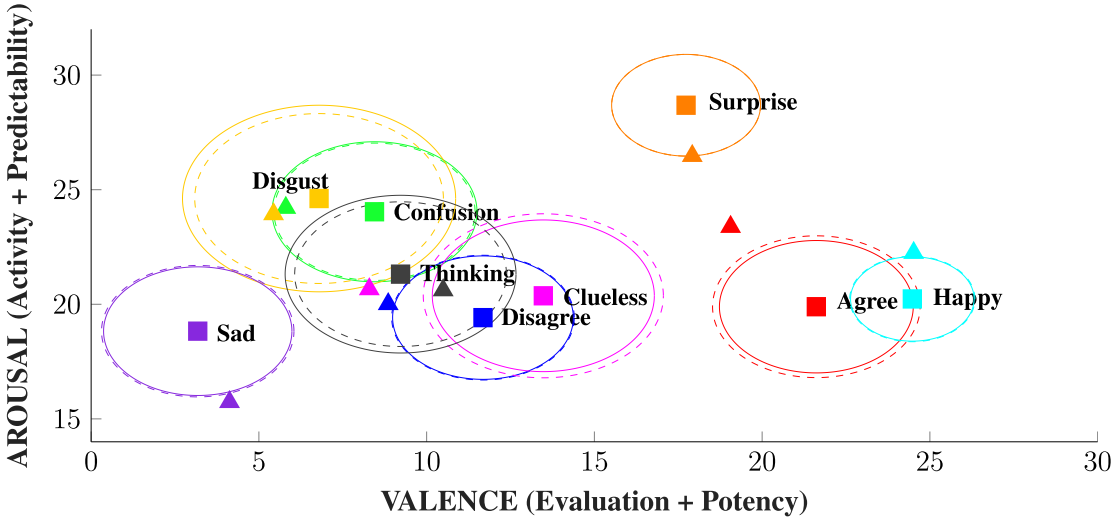

Figure 3.5: Projected scores of CJCm expressions (triangles) in our previous Semantic Space for videos available in Figure 2.7 (squares). Equivalent expressions are color-coded. The radii of the ellipses indicate the mean of the Euclidean distances of each actor to the expression.

As expected, the positions of CJCm differ from the averaged expression. Clearly, however, his data is very similar to that of the other actors: His data does not alter the mean Euclidean distances of the original set of actors by much (in some case it evens reduces them. A comparison of his values to the exact locations of the other six actors shows that there was always at least one actor in the MPI dataset whose performance was similar. Critically, the means show that the differences from CJCm to the average expressions are not due to the different information present in the videos or to cultural differences (the original videos are from German actors, ours are from Spanish actors). The differences of out actor are mostly likely due to merely to personality traits. It is also worth noting every actor is bad at, at least, one of these nine expressions [10], which might explain the one or two cases where the Euclidean means increased somewhat.

A different way of examining the data is provided using a Procrustes analysis to compare the position of CJCm's nine expressions in the space recovered in Experiment 3 to the Semantic Space given in Section 2.4. The standard distance measure is the Sum of Squared Errors between the two matrices, which in this case yields a distance $d = .303$. Since the correlation of the two matrices can be calculated from the SSq ($r^2 = 1 - SSq$) [45], we see that the two matrices are correlated at $r = .8348$ (the two are significantly correlated, $p < 0.001$). In sum, the Semantic Space recovered here is highly similar to that found in the previous chapter and the location of the new actor's expressions in that space is very similar to other actors. It is very likely, then, that the position of his other expressions is equally representative.

3.4 Experiment 4: Motion Capture Data

Many character animations rely on Motion Capture. Here we extract the Semantic Space using Motion Capture data for the exact same recording sessions used in Experiment 3. This will not only show us the perceptual relationship between different expressions in the Motion Capture but will also allow us to directly compare the Semantic Spaces for real video space and Motion Capture.

3.4.1 Methods

A total of 10 people participated (age range $23 - 33$; 5 females). The mean time to complete the experiment was 1 hour 48 minutes. The stimuli for this experiment were 62 rendered point-cloud videos (see Figure 3.3 for a snapshot of the point cloud) from the Motion Capture data corresponding to the 62 expressions shown in Table 3.1 where the same actor was showing one expression.

3.4.2 Recovering the Semantic Space

The parallel analysis, optimal coordinates and acceleration factor criteria agree that two factors are needed while Kaiser criterion, scree test, and explained variance suggest a three-factor solution is needed. We will explore both options.

In the 2D solution we obtain the factor loadings shown in Table 3.4, with Factors 1 and 2 explaining 48.8% and 24.4% of the variance, respectively, for a total of 73.2% (see Figure 3.6). The factor analysis successfully recovers the typical fusion of the EPAP dimensions in the form of Valence (Evaluation and Potency) and Arousal (Activity and Predictability) from the scales (all cells in Table 3.4 which are bold-font).

The three dimensional solution without any rotation (explaining a total of 82.21% of the variance), is the same as for the videos. The first factor is Valence (scales 1, 2, 5, 6, 9 and 10), explaining 44.1% of the variance, the second is Arousal (scales 4, 7, 8, 11 and 12), explaining 28.5% and the third factor (scale 3) would stand for 9.5%. When applying varimax rotation, the grouping of the scales into factors considerably changes: it fuses Evaluation and Potency into a Valence factor and keeps Activity and Predictability separated. Specifically, the second factor explains 20.1% of the variance and groups scales 3, 7 and 11, the first and third factors explain 47.3% and 14.8% of the variance respectively and the corresponding groupings are (1, 2, 5, 6, 9 and 10) and (4, 8 and 12).

Scale ID	Factor 1	Factor 2
1	**0.905**	-0.169
2	**0.920**	-0.114
3	-0.503	**0.558**
4	0.077	**0.886**
5	**0.968**	-0.011
6	**0.913**	-0.090
7	0.347	**0.660**
8	0.066	**0.446**
9	**0.945**	-0.097
10	**0.944**	0.037
11	0.392	**0.593**
12	0.309	**0.888**

Table 3.4: Factor loadings for the 2D space. The significant contributions recovered from our solution are shown in bold.

3.4.3 Comparison to Experiment 3

In order to see the viability of using the Motion Capture data space interchangeably with the real video space and, thereby being able to animate an avatar with the desired

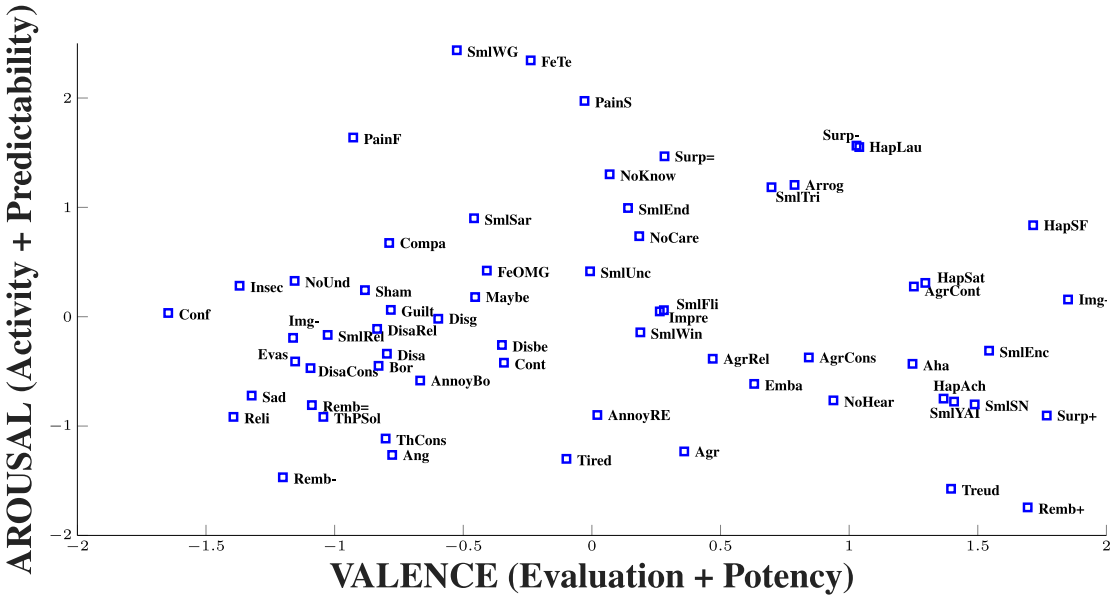

Figure 3.6: Coordinates along Valence (Factor 1) and Arousal (Factor 2) for the 62 Motion Capture videos.

MoCap-Data for an expression ensuring that that expression will be recognized as intended, we need to study the correlation between both spaces.

Towards that goal, we compare the location of each conversational facial expression in the 2D space recovered from real videos (see Figure 3.4) to the one recovered from the MoCap-drive point clouds (see Figure 3.6) using a Procrustes analysis. We find a distance $d = .6635$ and correlation at $r = .5801$ (the spaces are significantly correlated, $p < .001$), which is acceptable, but a bit low. The transformation required to align the two spaces consists of a 22 degree rotation and a scaling of $s = .58$.

The fact that the correlation between CJCm's videos and the MPI database videos was very high $r = .8348$ suggests that CJCm's expressions are not usual. The fact that the stimuli for Experiment 3 and 4 were from the exact same recording sessions (i.e., they are different recording-modalities of the same event) means that the only possible cause for the deviation between the space found in Experiment 3 and Experiment 4 must be due to the change of modality. In particular, for the Motion Capture videos out of plane rotations are harder to be detected (due to the lack of body as a reference-frame), there is a simplification of the movement (texture deformation such as wrinkles are not present) and, most importantly, the eye-movements are absent (the glance of the actors

was not recorded). Likewise, any body motion (such as shrugging the shoulders) will also be absent in the point cloud. These last two factors have been shown to be necessary for different expressions such as thinking [10, 53, 54].

To analyze the impact of the absence of gaze-tracking in our recording, we re-ran the analysis without those expressions that previous work has described as mainly driven by eye-motions (Thinking, Sadness and Clueless and all their subordinate-expressions [10, 53, 54]) and we observe an increase in the correlation ($r = 0.669$, $p < 0.001$, $d = .5524$, $rot = 14.24°$, $s = .68$). Since the necessary and sufficient facial motions for the majority of our 62 expressions have not been empirically established, there is no objective ground upon which a removal could be justified. Yet, a casual glance at the video shows that there are other expressions which use eye or body motion. It is very likely that if the comparison were limited to expressions that do not rely on the eyes or body, then the correlation would increase further. To empirically decide whether the eyes are necessary for each of the 62 expressions in our dataset, we would need to perform an extensive set of experiments to the one performed by Nusseck et al. [54]. Even though such a study is beyond the scope of this work, there are clear indicatives that it would be really advantageous.

3.5 General Conclusions

In the experiment described in Section 3.3, we recovered the Semantic Space underlying a set of 62 videos from a single person. The data can be well explained by either a 2D or a 3D space, and in both cases the space is consistent with previous work. The 2D is a classic Valence/Arousal space, while the 3D space is the standard EPA. Since moving from 2D to 3D only explains an additional 5% of the variance, we suggest that the more compact 2D space is preferable.

We then compared the performance on a subset of the expressions in this experiment to the performance on the same expression in the experiment described in Section 2.4 from the previous chapter. We found a high correlation between the current and previous spaces as well as the location of the expressions in those spaces. This leads to three conclusions. First, we can assume that, as the Semantic Space for facial videos is highly correlated with the Semantic Space derived for words, as presented in Chapter 2, there will be a high correlation between our videos and the words. That is, the new

expressions were correctly recognized. Second, the effect of culture (German Actors in the previous work versus the Spanish Actors in the preset work) is within the variance found for individual differences. Third, we found that we can use the loadings presented in Chapter 2 to project the perception of our new video into that previously recovered space. That is, it is easy to project any new recordings into the existing space without having to re-measure the ratings for all the previous videos. These findings allow one to expand the space using between participants designs, with different groups of participants rating subsets of the larger database.

In the second experiment described in Section 3.4, we used the same methodology to recover the space underlying the MoCap-Data corresponding to the videos of the previous experiment. We analyze the 2D (Valence/Arousal) and 3D (Valence/Activity/Predictability) solutions. Just as with Experiment 3, the 2D space was preferable. We then compared the location of specific events (a given instance of an expression) as recorded either by video or by MoCap). The location of videos of expressions significantly correlated to the MoCap recordings. Removing expressions that are known to rely on eye motion (which was not recorded in the MoCap recordings) improved the correlation. This shows that the MoCap-Data is missing information that is critical to the perception of facial expressions. Any animation that relay solely on MoCap will be incomplete and possible misperceived. This clearly shows that when creating character animation, MoCap-Data need to be augmented with additional modalities such as eye tracking to more accurately reflect human socio-emotional behavior.

As already pointed out in the previous chapter, the deviations from the mean of one actor can give some insights on his personality profile. Given that our database contains a very large number of expressions, it is viable to perform an experiment were people would measure the perceived personality from the actor by filling an standard, validated questionnaire such as the Five-Factor Model Rating Form (FFMRF) [55], which we do in Chapter 6. This would allow us to map the personality space to the facial expressions one, enabling style motion transfer. Finally, the technique can be used to empirically map specific motion trajectories (or their frequency-decomposition) onto specific perceptual attributes, allowing the targeted creation of novel animations with the desired perceptual traits, which is done to some extent in Chapter 4.

Part III

Motion Synthesis

In this part, we provide a proof of concept that the metric mapping from facial expressions to the Semantic Space can be inverted to generate new facial expressions. We select a new location in the Semantic Space, by providing its Valence and Arousal coordinates, and then use a weighted combination of the Motion Capture recordings, with the weights being determined using a distance metric in the Semantic Space. To test the accuracy of the method, we perform a leave-one-out analysis.

DERIVING EXPRESSIONS FROM THE SEMANTIC SPACE

4

Here, we use the recovered Semantic Space along with the mapping to the Motion Capture recordings obtained in Part II to synthesize a novel facial expression with the desired emotional tone. Specifically, we select a new location in the Semantic Space, by providing its Valence and Arousal coordinates and then calculate the distance from that point to all known locations in the Semantic Space. To create the new animation, we blend the Motion Capture recordings that correspond to the known locations, using the distance in Semantic Space to the new point to define the weights. To test the accuracy of the method, we perform a leave-one-out analysis. That is, we select as coordinates for the new animation a point in space that corresponds to a known expression. We then use all the *other* recordings to create an animation for that point. We then compare the reconstructed animation to the original for that point (which, of course, was not used in the reconstruction).

The content of this work is yet to be published and was done in cooperation with Katharina Legde (Graphic Systems Department, BTU Cottbus-Senftenberg), who helped with the cleaning and labeling of the Data Base and the advisor of this thesis, Prof. Dr. Douglas W. Cunningham (Graphic Systems Department, BTU Cottbus-Senftenberg) who oversaw the project.

4.1 Introduction

In the previous chapters we used psychological methods to recover a vector space describing the semantic properties that humans perceive in facial expressions. We showed that the same space can be used to describe the perceptual properties of words describing expressions, videos of facial expressions, and Motion Capture point clouds. The Semantic Space is incredibly robust and has many interesting properties, seeming to capture an amazing range of perceptual attributes. The previous work recovering this Semantic Space also provides a one-to-one mapping between individual recordings and locations in this space. The mapping is sufficiently stable that one can use it to project new facial expressions into the space, as we did in Section 3.3.3, to learn about the relationship between these new expressions and the previously tested ones. In principle, it should be possible to "invert" the projection, and go from a location in the Semantic Space to create a new facial expression.

The core to inverting the mapping relies not just on the fact that we now have a one-to-one mapping from facial deformations to semantic meaning, but on the fact that Semantic Space and the facial deformations are in coherent vectors spaces. We have already shown, in detail, that the Semantic Space is a Hilbert space. The representation of the facial expressions is however, still open. While one can debate about the best way to represent the information in the Fontaine's emotional words or in the video recordings, the best way to represent the Motion Capture data us clear. In 1999, Volker Blanz and Thomas Vetter introduced the concept of the Face Space [56]. The scanned in 200 human faces, producing a set of sample locations on the faces (vertices) which were connected in a mesh. They re-meshed all the scans so that they had the same number of vertices (approximately $n = 70.000$) each of which was located at the same point on the face. They then converted the spatial coordinates of each vertex (X,Y,Z) into line vectors, and concatenated the vectors together to make a $3n$-dimensional shape-vector defining the geometry of the face. Similarly, they also defined a $3n$-dimensional texture-vector containing the color values (R, G, B) of the n corresponding vertices. These two vectors defined the "face vector". Each of the 200 scans, then, are considered to be a double $3n$-dimensional vector in the same space, and as such weighted combinations of them can be used to produce new 3D meshes of faces. This is the "Morphable Face Model". They subsequently showed that if the individual face vectors where labeled (e.g., such as being from a male or female, as being happy or sad, etc.), then the face-space vectors

could be combined in ways that took advantage of specific (perceptual) properties. For example, they found the average face-vector over all 200 faces and then the average male and average female face. The difference between the average male face-vector and the average face-vector can be seen as the bias in a face that is due to being female, on average. They then subtracted the "male" bias vector from a given male face and added the average female bias vector to create that person's twin sister. Given a labeled dataset, the possibilities for combinations are endless.

In our case, mapping from the Motion Capture videos to the Semantic Space provides us not only with a labeled database, but the labels are themselves in a vector space. Since all the Motion Capture recordings were made with the same marker setup, the spatial sampling of all the expressions for all actors are already in correspondence. Thus, we can convert each Motion Capture recording into a vector containing the spatial coordinates of each of the vertices. The main difficulty is that we do not have static faces, as Blanz and Vetter did, but dynamic faces. That is, each vertex has a series of spatial locations. Unfortunately, the number of frames (and thus the number of spatial coordinates for a given vertex) differs across recordings, so that a simple application of the morphable face model concept to our data is not possible. While one could, in principle, merely use the neutral facial pose and the peak for a given expression, and then morph linearly between them in time to create an animation, previous work on the perception of facial expressions (e.g., [47, 57, 58, 59, 60]) has shown that temporal information is critical to the proper perception of facial expressions. The other alternative is to re-sample the frames somehow. The most obvious solution is to place the motion trajectories into correspondence, such as using Martin Giese's techniques [61]. These techniques were, however, designed for rigid body motion and their application to facial motion is very non-trivial. Moreover, they tend to change the acceleration and velocity profiles of the motion, which will alter the perception of facial expressions [57]. As Nikolaus F. Troje has shown in his work on morphing rigid body motion, a very good first approximation of motion blend can be achieved if one merely performs a frame-by-frame combination, assuming that the starting points of the two sequences are more or less synchronized [62].

In this chapter, we provide a simple technique to show the generative capabilities of the Semantic Space. It relies on the assumption that the Motion Capture recordings can be represented as a temporal series of face-vectors whose starting points are more-or-less synchronized, and that the mapping to the Semantic Space converts these face-vector

series into a labeled database. To create a new animation with targeted perceptual properties, we select a point in Semantic Space, find the distance to all other points in that space, and then use these distances as the weights to combine the corresponding face-vectors. We combine temporal sequences on a frame-by-frame basis, following the work of Troje [62].

4.2 Data Pre-processing

In order to use the MoCap recordings, the data must first be pre-processed. Since raw MoCap recordings tend to do be rather noisy, the first step is to clean the data. This involves five main steps. First, during the recording process some markers get lost (such as when they are temporarily not visible). When they re-emerge, they are considered to be a new marker. Thus, the first step is to merging markers that were lost and then re-appeared with different labels. To do this, we find all markers that were always visible during the full recording. These will provide a stable face-specific framework for recovering the remaining markers. Then, we find the markers that have temporal discontinuities, and calculate the distances of each of these "disappearing" markers to the always visible ones. Disappearing markers with similar distance matrices are candidates for merging. We merged those presenting the minimum variance on these distances along time, with a 95% of confidence. Next, we fill in all gaps in all the motion trajectories. To do this, we find the three closest neighbor markers to the maker for which we wish to fill the gap, use a Single Value Decomposition (SVD) to calculate the average translation and rotation of these three neighbors and transferred this movement to the lost marker to generate its trajectory for the empty frames.

Having completed all motion trajectories, the next step in cleaning is to remove undesired high-frequency jitter present in the original trajectories. To do this, we used a butter filter (band-width-filter) on the full recovered sign. This also helps to remove any possible artifacts that may have been introduced in the filling-gaps process.

Subsequently, we remove the rigid head motion via SVD (using the markers in the head and ears). The rigid head motion was stored then in a marker "RHM" containing its x- y- z-displacements and its Euler rotations.

Next, we label each marker. Although the initial attempts at cleaning had the labeling done by hand using the static recording of the actor as a reference, we later replaced this with a Voronoi-Diagram-based technique using the first frame of each expression (since we know that all markers will be visible there). We take the point cloud, create the corresponding Voronoi Diagram for it, and then compared the diagram with a standard, labeled Voronoi Diagram. In order to assure the robustness of the labeling, this process was repeated each 300 markers to prove its coherence. In case the labeling process gave different labels for the same marker depending on the frame used, we selected those labellings that had the highest number of hits on a given cell among all repetitions of the process. In order to align the mask with the frame, we used SVD on the position of the head markers of both the static mask and the frame to align them.

Finally, we execute a manual review of the final results and manual corrections. One of the most common errors included removing double markers on the eye-lids (caused by blinking).

After the database was cleaned, it needed to be parsed. As each recording corresponded to three repetitions of the expression (for more on how the recordings were made, please see Chapter 3). The actor returned to a neutral expression between repetitions of the reaction. The best recorded repetition for each expression was manually selected from the video recordings and, as the Motion Capture data was synchronized with these recordings, we apply these same start and end frames to the MoCap recording to obtain the best repetition (for more on the parsing of the video recordings, please see Chapter 3). Please note that finding the so-called "true" start and stopping points of each expression within a recording is not a trivial issue, as it is quite a subjective task. The criteria used to establish the beginning of the expressions was, while visualizing the video recording in a frame-by-frame manner, we established as start point the previous frame to the one where, being the actor in a neutral position, we visually detected any sight of muscular activity in any facial area that the actor showed when starting to show the expression. Analogously, the selected frame for the ending point of the expression was the one directly after the actor has show the last visually perceptible muscular relaxation in the face to go back to the neutral expression.

4.3 ALGORITHM

In our algorithm, we use the previously established mapping from MoCap to Semantic Space along with the cleaned and parsed Motion Capture recordings to synthesize new expressions. The first step is to convert each Motion Capture recording into a facial expression matrix, which encodes the marker name, the frame number, and the 6 degrees of freedom representing the point (x, y, z, rotation along X, rotation along Y, rotation along Z). The full set of facial expression matrices is our Motion Capture *dictionary*. The entries in the dictionary are converted into a form more suited to blending. Specifically, we convert the spatial position of each marker on all but the first frame into a spatial displacement of the marker from its position on the previous frame. We then create the label space dictionary, which contains the coordinates of each recording in the recovered SSp (see Chapters 3 and 6), along with the name of the actor and expression present in that recording. The combination of Motion Capture Dictionary and Label Space Dictionary can be used to create new animations.

To create a new animation, a new location in the Semantic Space is inputed. The distances between this new point and all known points in the Semantic Space are calculated (using the Label Space Dictionary), and stored in a distance matrix. These distances will be used as the weights for combining the relevant entries in the Motion Capture Dictionary. Note that any number of distance functions are, in principle, possible. For the purposes of this proof of concept, we test the two most common forms: the inverse and the inverse squared of the Euclidean distances. As is common in this form of scattered data interpolation, the weights are normalized to be between 0 and 1, and to sum up to 1. The last step of the preparation is to calculate the desired duration of the animation. Since we are using a frame-by-frame blending, this is the maximal length of the two recordings. If one sequence is longer than the other, the shorter sequence will need to have additional data added so that the two facial expression matrices have the same dimensionality. There are several options here, the simplest of which is to simply hold the last known position. That is, all markers in all subsequent frames will have no displacements from their last location, and as such the new values in the facial expression matrix are all 0. One could also repeat the expression. Since all expressions recordings start with a neutral expression and end with a neutral expressions, adding a repetition of the expression by coping the values from the second frame onwards should be possible without adding noticeable artifacts. Finally, the all facial expression matrices in the

Motion Capture Dictionary are blended using the weights in the distance matrix to create the new Motion Capture sequences. As an additional touch, one can add back the (also blended) rigid head motion.

To visualize the Motion Capture data, one will need to use to drive some form of animation. The simplest technique would be to use the motion to drive a set of cubes, creating a point cloud animation. Alternately, one could use the motion to manipulate animation rigs on a facial mesh.

4.4 Results

To test our motion synthesis technique, we used our Motion Capture recordings from one actor (CJCm) and the SSp recovered for that actor (see Chapter 3). We then used point cloud animations to visualize the motion. We tested the synthesis with a leave-one-out analysis. That is, we took a known location and expression, such as Happy, and removed it from the Motion Capture Dictionary and the Label Dictionary. We then inputed the coordinates of the removed expression (e.g., Happy) to synthesize it using all the other expressions. We can then compare the re-created expression with the original (recorded) one. For example frames, see Figure 4.1.

Overall, the results were quite convincing and easily recognizable as the intended expression. Initial tests suggest that the inverse squared distance function produced better results. Some expressions were easier to synthesize than others. For example, the subordinate-emotional expressions belonging to the family of Smile (see Appendix A) have quite a resemblance to the originals, while Angry, for example did not. This might be related to the way the actor we synthesize expressed some emotions. In the case of Angry, the actor stared at the interlocutor, and slowly titled his head towards one side while the veins in his neck became clearly visible, logically, this translated into just a rotation in the head in the Motion Capture file.

It is also worth to mention that, when synthesizing combined emotions (e.g. Pleasant Surprise or Agree Consider) the nature of the expressions was critical for their closeness to the original. Combined expressions which consisted on the simultaneous occurrence of two or more basic expressions, as the case of Pleasant Surprise (when a person is happy and surprised at the same time), can be expressed as a weighted combination of emotions. On

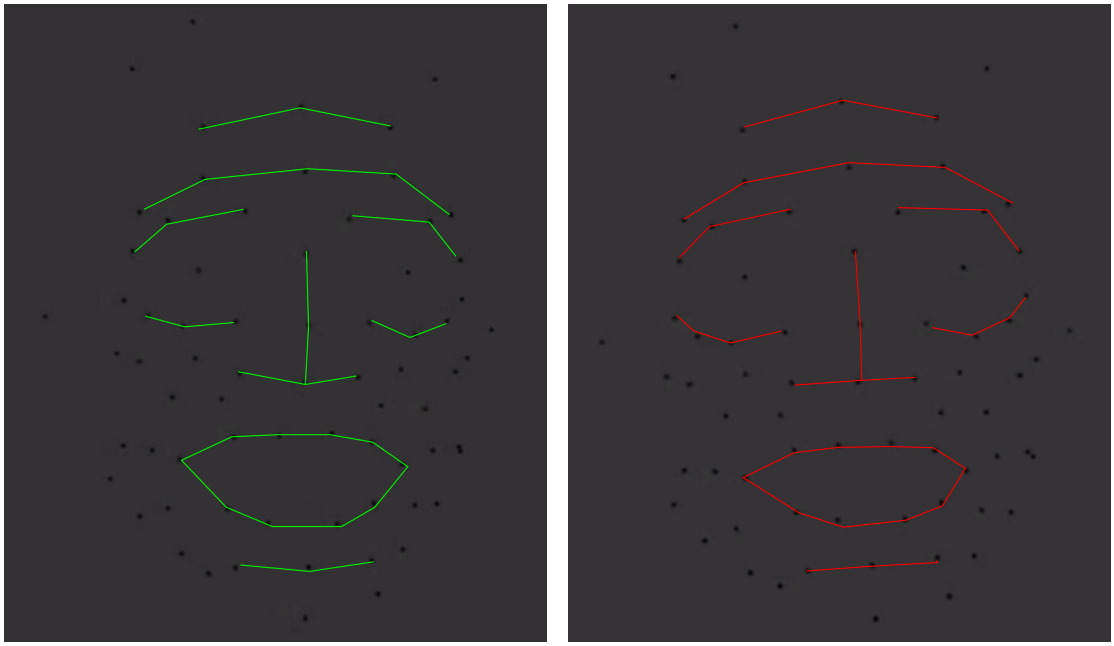

Figure 4.1: Snapshots for the expression Happy ("HappyLaugh") both in its original recorded form (left), and the result of the proposed technique (right). For better visualization, some guidelines have been added to the point cloud.

the other hand, combined emotions that consist of two or more concatenated expressions, such as the recorded "Agree Considered", will not be comparable to the weighted combination of its parts. For the mentioned example, what is recorded is a person who first considers an idea, and then agrees to it, which is intrinsically different to a person showing agreement (nodding) while also considering something.

4.5 Conclusions and Future Work

We believe that, as the use of such a simple technique gives us already quite acceptable results, it illustrates how powerful the generative capabilities of the Semantic Space can be. The code allows one to easily change the Semantic Spaces to consider or the distance function to use. It is also straightforward to modify the weights of the expressions to consider, not only their distance to the point to generate, but also extra factors such as clustering-expression for subordinate emotional expressions (see Appendix A), or quadrants' weights. Moreover, manipulating the algorithm to consider, for example

personality deviations (given their mapping to the Semantic Space), is extremely easy. This technique, together with the findings on the following part, offers a promising research platform for the study of facial communication in a controlled, systematic fashion.

Our results have proven that recovered dimensions of the Semantic Space are related to the facial actual movements, but the correspondences between the trajectories of the markers that form an expression and the coordinates of the latest on this space is not trivial. In future work, we aim to find the structures that define specific emotions, in order to provide an empirical basis for determining which aspects of the videos are important and which aspect of the expressions' meaning they might carry. This would allows us to use more refined combination of recording elements in order to produce novel expressions. Thus, we will avoid using all motion in all of the recordings but rather only the relevant, meaning-carrying motions. In other words, by exploring the trajectories of the emotions sharing one dimension, either by clustering significant areas of the face (mouth, eye-browns, etc.) or simply by studying overall properties of the movements corresponding to the expressions (e.g. intensity, duration – peaks and plateaus –, and rapidness –acceleration) we expect to be able to characterize the dimensions of the Semantic Space. This would make possible, given its coordinates, to reproduce or generate any expression in the Semantic Space as one would know which motion parameters should be used (i.e which facial markers' should move, in which directions, how much, and which velocity and acceleration profiles should be used).

In this direction, we have already conducted a preliminary study on the markers' trajectories of the recorded expressions. The results from this exploratory work indicate that a high value on Arousal could be related to a short time for the expression to reach its peak (i.e. high acceleration profile) and also a short plateau – time holding the peak of the expression. Consequently, a low Arousal value will correspond to expressions with a long transition from neutral to peak and long sustain. Also towards parameterizing the animations and, therefore, allowing us to avoid using specific recordings, we have also begun to look for systematic patterns of motion for collections of markers across recordings, finding some high-order physical structures (eye motion, mouth motion) in the MoCap recordings. Analyzing the data by clusters inside quadrants, gave us some insights on the most important markers for each cluster and indicated a dependency between the horizontal and vertical displacements of these markers and the Valence of the expression.

Part IV

Personality

This part explores the correlation between the facial expressions people use to convey an emotion and their personality. We discuss how the shift between one particular person's expression to the averaged expression is an indicator of their personality. Thanks to this correlation, we can provide Embodied Conversational Agents with individual personalities by shifting their facial expressions according to a desired personality profile.

PERSONALITY ANALYSIS OF ECAs

5

People tend to personify machines. Giving machines the ability to actually produce social information can help improve human-machine interactions. An *Embodied Conversational Agent* (ECA), as mentioned in Chapter 1, is a virtual software agent that can process and produce speech, facial expressions, gestures and eye gaze, enabling natural, multimodal, human-machine communication. On the one hand, the field of personality psychology provides insights into how we could describe and measure the virtual personality of ECAs. On the other hand, ECAs provide a method to systematically examine how different factors affect the perception of personality. This chapter shows that standardized, validated personality questionnaires can be used to evaluate ECAs psychologically, and that state of the art ECAs can manipulate their perceived personality through appearance and behavior.

An edited version of this work is published in the *Proceedings of the ACM International Conference on Intelligent Virtual Agents* and was presented at the *18th ACM International Conference on Intelligent Virtual Agents*. The co-authors of the mentioned paper were Philipp Hahn and Katharina Legde (Graphic Systems Department, BTU Cottbus-Senftenberg), who helped with the stimuli generation and analysis of the results, and the advisor of this thesis, Prof. Dr. Douglas W. Cunningham (Graphic Systems Department, BTU Cottbus-Senftenberg) who oversaw/guided the project.

S. Castillo, P. Hahn, K. Legde and D. W. Cunningham
Personality Analysis of Embodied Conversational Agents
In Proceedings of the ACM International Conference on Intelligent Virtual Agents, IVA 2018

5.1 Introduction

Life is inconceivable without communication, which occurs vocally as well as through face and body movements. While communicating, we do not only convey the semantic content of the message but a lot of other information – such as our emotions, mood or affective ties [48] – often subsumed under the name socio-emotional content. Given our extensive experience with communication, it should not be surprising that we are experts at both production and interpretation.

Although machines generally neither process nor produce socio-emotional information, people still tend to base their interaction with machines on apparent interpersonal behavior cues coming from the machine [49, 50]. The field of Affective Interfaces tries to improve human-machine communication by explicitly giving computers the ability to process and/or produce socio-emotional information (see, e.g. the work of Cassell et al. [63]). For example, ECAs present virtual characters who are able to interact with human beings by interpreting and producing multimodal communicative behavior. Of course, as soon as computers are given a virtual body and/or voice, it becomes even more important to explicitly address the virtual socio-emotional signals. Since computers do not really have emotions or intentions, let alone personalities, ECAs – and other affective interfaces – must explicitly model the computer's intended personality and map that to the computer's behavior.

Personality is a term used to describe stable qualities of how an individual acts and reacts. It reflects the person's characteristic behaviors, emotions, intentions, wishes, and values [64]. A person's personality is a strong determinant of how others act or react to him or her, including, for example, how tolerant they are to mistakes. The description and measurement of personalities is a huge research field within Psychology, with a long history. Although personalities are very complex, it is generally accepted that there are some broad commonalities among all people. The most accepted descriptions of personality state that all people can be rated along several basic dimensions, and this provides a rough but more-or-less complete description of their basic personality.

Naturally there are different opinions about the number of dimensions needed to describe a personality. There are simple approaches like the two dimensional model proposed by Hans J. Eysenck [65], where only Extroversion and Neuroticism are used. At the other extreme are models like Cattell's model [66] which is based on 16 factors that the author

suspected to have more than five underlying dimensions [67]. Perhaps the most dominant personality model – within Psychology – is the Five-Factor Model [68, 69, 70]: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism which is referred to as OCEAN or the Big Five. Each of the five main dimensions is a conglomerate of related but slightly different traits that are useful in measuring the full spectrum of personality. For more on OCEAN, please see Section 5.2.2.

Is worth mentioning that the reference corpora for the standardized tests used to evaluate personality are based on human data. Therefore, it could be questionable if they can be used on ECAs. Nevertheless, giving that previous research has proven the fundamentally social nature of the interaction between humans and computers [71], it is not unreasonable to assume that computers can have human-like personalities and, thus, these can be measured with the same tools as employed on humans. Moreover, the State-of-the-Art report of Vinayagamoorthy et al. [72] indicates that people tend to personify computers and hence are likely to respond to them in the same manner as they would to real humans. Furthermore, they point out that people are able to identify the intended personalities of virtual agents. In addition, Cafaro et al. [73] concluded that the findings on personality assessment derived from social psychology research are valid and applicable to virtual agents.

The design and evaluation of the behavior of most ECAs (e.g. [74, 75, 76, 77, 78]) has used either simple models or focused on a small subset of the more complex models. Although these simplified representations can satisfy the generic needs to design a character's personality, they also can be insufficient to seize its specific nuances. Moreover, by not looking at the whole of personality, unattended dimensions may end up with undesired (and undesirable) values. Here, we propose that OCEAN is just as useful for evaluating virtual personalities as it is for human personalities. Moreover, we hypothesize that even a simple approach to measuring OCEAN is sensitive enough to pick up the subtle differences in personality. If true, then the combination of ECAs with the OCEAN model holds considerable promise for both Psychology and Affective Interface research. Since ECAs allow us to systematically and carefully control (behavioral) changes, we would then be able to create a theoretically driven, empirically obtained mapping of personality to behavior (and back). The first step, however, is to show that a simple questionnaire for measuring OCEAN is indeed sensitive enough to pick up both large-scale and subtle changes in virtual behavior, and to provide some initial insights into the perception of personality. The central research questions of this chapter are:

- Can an OCEAN questionnaire measure the perception of an ECA's personality?
- Does the perceived personality fit the designers intentions or are there undesired traits?
- How do auditory and visual information contribute to the perception of an ECA's personality?

5.2 Evaluation Framework

For our experiments we used a standardized personality questionnaire (see Section 5.2.2) on a State-of-the-Art ECA (see Section 5.2.1).

5.2.1 Sensitive Artificial Listener (SAL)

The SEMAINE API [79] (Sustained Emotionally colored MAchine-human Interaction using Non-verbal Expression) is designed to be a robust, real-time system capable of analyzing and synthesizing multimodal behavior. In this chapter, we use a subsystem of the SEMAINE API 3.1, the Sensitive Artificial Listener (SAL) [80, 81], which provides as potential virtual dialog partners four different avatars, each with its own personality, voice and appearance (see Figure 5.1). *Spike* was designed as an aggressive, argumentative character. *Poppy* is supposed to be outgoing and cheerful. *Obadiah* is meant to be pessimistic and gloomy. *Prudence* should be pragmatic and reliable [82].

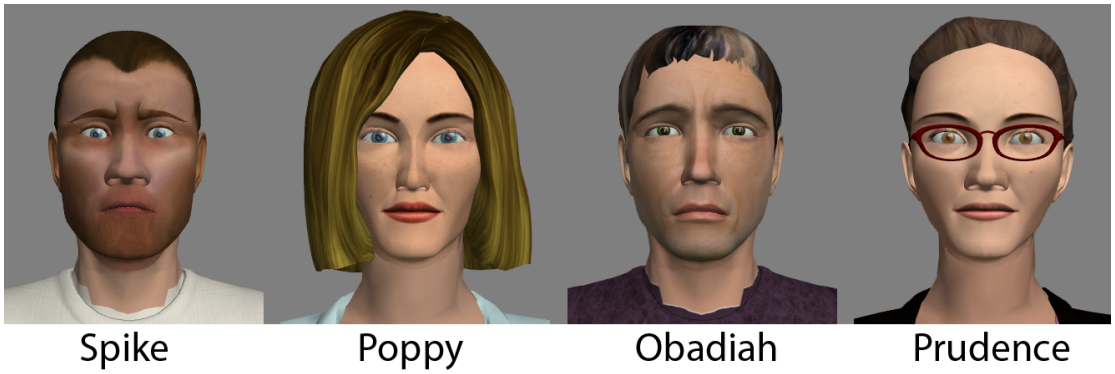

Figure 5.1: SAL system - The four different avatars.

It is worth emphasizing that SAL is an artificial listener and the avatars have the explicit goal of encouraging the user to continue talking. As such, SAL's behavior is focused on so-called *back-channel* signals, which refer to the exchange of signals from the listener(s) to the speaker [17]. Thus, SAL does not need to understand everything that is said [80, 83]. Furthermore, the actual sentences used by the avatars are semantically quite reduced, although they do contain information about the mood and personality of the avatar. Indeed, most of this information could – in principle – be communicated solely through facial expressions and simple sounds. Here are some of the sentences used by each of the avatars:

Obadiah: *Things often get worse. / There is not much you can do about it. / Just think about all those depressive things. / I am not so sure you should be so neutral about it.*

Poppy: *That sounds interesting, tell me more about it. / It is great to hear someone sound so happy. / I think you have done really well. / I am glad to hear that.*

Prudence: *You obviously have your head screwed on. / I admire that. / Tell me what is going on at the moment. / What do you think it will happen?*

Spike: *What is your problem? / You are so pragmatic. / You sound like an airhead. / Life is a war, either you are a winner or a loser.*

In order to provide the avatars with their intended personalities, the designers placed them on the two dimensional personality model of Eysenck [65] (shown in gray in Figure 5.2). They then, based on previous research, mapped the Neuroticism value to the type of back-channel behavior and Extroversion to the back-channel frequency. Specifically neurotic characters such as *Spike* and *Obadiah* will tend to have more reactive responses whereas the more emotionally stable *Poppy* and *Prudence* will tend more towards mimicry. Likewise, the introverted *Prudence* and *Obadiah* will have lower overall activity levels than the more extroverted *Poppy* and *Spike* [82, 84, 85]. In an initial evaluation, the designers found that the desired type of Extroversion was perceived for outgoing and pragmatic personalities and that the intended degree of Neuroticism was perceived correctly only for pessimistic personalities [86].

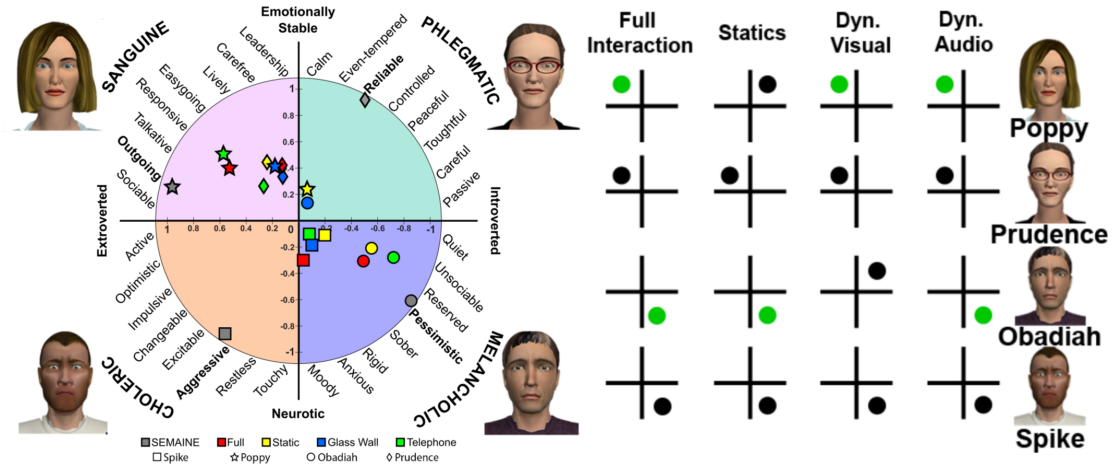

Figure 5.2: Coordinates of the four avatars in Eysenck's 2D personality space for the intended positions (in gray) from the SAL's designers and the resulting positions for the averaged E and N dimensions extracted from the FFMRFs of our experiments (different colors per scenario). On the right of the image, there is a simplified visualization of the quadrants where each SAL's avatar was located in each experiment.

5.2.2 Five-Factor Model Rating Form

Nearly all methods for evaluating an individual's personality involve directly asking someone some questions. Previous work in evaluating the personality of virtual characters has focused on a subset of the OCEAN dimensions and has found interesting insights (e.g., [74]). Unfortunately, creating a proper questionnaire, and ensuring that it is (simultaneously) externally valid, internally valid, and reliable is a very difficult, time-consuming procedure, requiring a lengthy, carefully controlled validation process [87]. Fortunately, there are a number of standardized, validated personality questionnaires. Some of them involve hundreds of questions – and allow a detailed examination of very subtle aspects of personality. The most widely used questionnaire, for example, is Costa and McCrae's NEO Personality Inventory - Revised (NEO-PI-R) [88], which has 240 questions. These same authors also showed that reducing the number of scales to a specific set of six per dimension was sufficient to determine their overall nature [88]. The questionnaire we used in our experiments, the FFMRF [55], is based on the NEO-PI-R. With a total of thirty 7-point Likert items, this standard validated form has 6 bipolar scales for each of the OCEAN dimensions. Each scale can get ratings inbetween 7, referring to an extremely high suitability of one of the anchoring descriptors for the specific scale (such as Pessimism) and 1 referring to an extremely low amount of it, thus,

matching the opposite anchor of the scale (Optimism). For an enumeration of the Five Factors and corresponding traits, please see Table 5.1, and for an example of the forms used in our experiments see Appendix B.

	FACTOR				
Sub-scale	**(O)penness**	**(C)onscien-tiousness**	**(E)xtroversion**	**(A)greeable-ness**	**(N)euroticism**
1	Fantasy	Competence	Warmth	Trust	Anxiety
2	Aesthetics	Order	Gregariousness	Straight-forwardness	Angry Hostility
3	Feelings	Dutifulness	Assertiveness	Altruism	Depression
4	Actions	Achievement	Activity	Compliance	Self-Consciousness
5	Ideas	Self-Discipline	Excitement-Seeking	Modesty	Impulsiveness
6	Values	Deliberation	Positive Emotions	Tender-Mindedness	Vulnerability

Table 5.1: The Five Factors of the OCEAN model and their corresponding scales.

5.3 General Methods

We performed three experiments measuring the effect of semiotic channels on the perception of personality. The first two experiments, Experiments 5 and 6, each had one semiotic or informational channel scenario. The last experiment, Experiment 7 had two scenarios. Each experiment followed the same general procedure. A total of 40 people participated (age range $21 - 33$), with each scenario using a different group of 10 new participants (5 females per group). All the participants were payed 8€ per hour for their participation. Each participant was informed of how the experiment would run – but not the research questions behind it – and was given a chance to ask questions. They were informed that they could stop the experiment at any point without any negative consequences to them. They then were asked to fill out a consent form.

The participant was sat in a semi-dark room in front of a 24" LED monitor (at a resolution of $1920x1080$) placed roughly 50 cm away and was equipped with a headset. The computer had the SEMAINE API 3.1 (using Set B for audio analysis [85]) installed. Since the version of the ECA we have only speaks English, the conversation was also in English, even though all participants spoke German as their native language. Only one participant performed the experiment at a time and was left alone in the room after the set-up was completed.

For all the experiments where some interaction between participant and avatar took place, the participants were given an explanation about what an ECA is and were given a few examples of conversational topics. They were asked to freely talk to the avatar and they were able to see its expressions and/or hear its answers, depending on the specific scenario. No text transcription of the conversation was present. All the experiments follow a within-participants design. After interacting with all four avatars for two minutes each (in a random order), the participants were asked to fill out the FFMRFs for the four avatars. The average time to complete each experiment was 30 minutes.

5.4 Experiment 5: Full interaction with an ECA

In the first experiment of this chapter, ten people interacted one at a time with the full system, including all input and output channels. Note the resemblance to a normal face-to-face (or a one-on-one video-conference) conversation.

The results clearly indicate that *standard personality measures can be used to evaluate virtual personality.* As can be seen in Figure 5.2 (plotted in red in the left image, and first column of the right image), when looking solely at Extroversion and Neuroticism dimensions of the FFMRF, *the SAL system was able to produce part of the desired personality profiles.* Although two of the avatars were placed on the intended side for both dimensions, the other two avatars were on the unintended side for one dimension. Specifically, *Spike* and *Prudence* were not in the intended half of the Extroversion dimension. This is quite different than the results found by de Sevin et al. [86], which may be attributed to the different personality measures used. Since we measured personality with the FFMRF, we are able to obtain a finer-grained examination of the extroverted and neurotic behavior of the avatars. Interestingly, Assertiveness (E3) and Excitement-Seeking (E5) placed the aggressive *Spike* exactly where he was supposed

to be and E5 also placed *Prudence* where she should have been (see the corresponding graphs in top left of Figure 5.3). This suggests that the *mapping of personality to behavior is true for a subset of each dimension, but in general needs to be much more complex.*

To examine the full personality spectrum, we calculated the average of the six sub-scales of each OCEAN dimension for each participant. Each dimension was then separately submitted to a one way ANOVA with avatar as a within-participants factor. Every dimension show a statistically significant variation across avatars (all $F's > 10.93$, $p's < 0.001$). In other words, as can be seen in the OCEAN graph on the top left of the first quadrant of Figure 5.3, *people thought that the different avatars had different personalities.* The aggressive *Spike* was significantly less agreeable than the other avatars (confirmed with two tailed t-tests, all $p's < 0.002$) and was also more neurotic than the pragmatic *Prudence* or the optimistic *Poppy* ($p's < 0.003$). Note that although the difference in Agreeableness matches the designers' goal implicitly, they did not explicitly model it. Both female characters were rated as being more extroverted ($p's < 0.003$) and less neurotic ($p's < 0.02$) than the depressive, gloomy *Obadiah.* The neutral *Prudence* was more open than both negative male characters ($p's < 0.01$) but less than the outgoing *Poppy* ($p < 0.05$). A closer examination of the sub-scales shows an interesting degree of complexity, some of which was clearly not in the design goals.

5.5 Experiment 6: Effect of Appearance

Of course, it is very possible that the differences in perceived personality do not come from different behaviors, but merely from the appearance of the avatars (i.e., the geometry and texture of the avatars' faces). Thus, in the second experiment of this chapter, Experiment 6, ten new participants were given photographs of all four avatars in a neutral pose (using a screen-capture of the avatar in waiting pose, with the mouth closed; see Figure 5.1) and were asked to rate that avatars using the FFRMF. Note that each participant saw the photographs one at a time in a random order – with each participant receiving a different random order – and did not interact with the avatars at all. The mean time to complete the experiment was slightly shorter than the one for the rest of the experiments, taking on average 23 minutes.

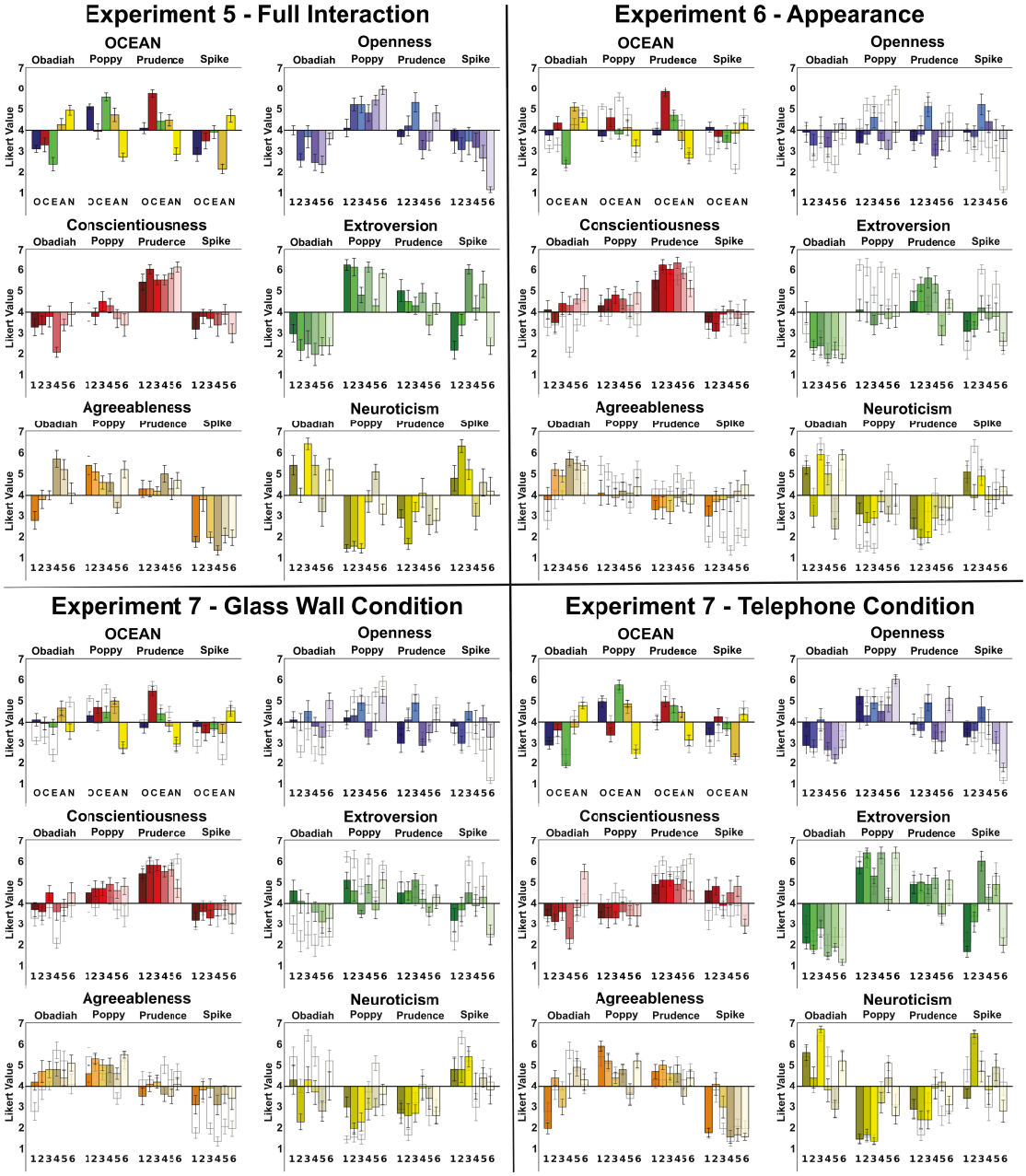

Figure 5.3: *Results of the 3 experiments.* For each experiment, on top left: Ratings for the Big Five Factors; Remaining plots: Scores for all 30 scales for each avatar (error bars represent the SEM). For Experiments 6 and 7, the grayscale silhouetted bars are the results for Experiment 5, so the differences can be inferred.

Clearly the different photographs were seen as belonging to different personalities. The corresponding one-way ANOVAs with avatar as a within-participants factor for each OCEAN dimension showed that all dimensions had a statistically significant variation across avatars with the exception of Openness ($F(3, 27) = 0.549$, $p > 0.65$; all others: $F's > 3.0$, $p's < 0.05$). Just as clearly, the static personality was often quite close to the one in the full system (see the left plot in Figure 5.2 and compare yellow versus red or refer to the corresponding graphs on top right of Figure 5.3 for the individual scales). *Obadiah*, *Spike* and *Prudence* were each in the same quadrants of the Extroversion-Neuroticism space in both experiments. *Poppy*, on the other hand, is in the unintended half of Extroversion in the static condition. Just as was the case in Experiment 5, when we only consider Excitement-Seeking (E5), *Prudence* would have been in the intended quadrant. For *Spike* and *Poppy* no sub-scales fit the desired Extroversion (but the ratings of the latter in all sub-traits were close to neutral; see top right of Figure 5.3).

To test the difference between the two experiments, we submitted the results of each OCEAN dimension to a two-way ANOVA with avatar as a within-participant factor and scenario (multimodal versus static) as a between-participants factor. The results showed that for all the five dimensions, there was an effect of character ($F's > 6.4, p's < 0.001$). There was no main effect of scenario ($F's > 0.08, p's > 0.2$)) but there was an interaction ($F's > 6.3, p's < 0.001$) between character and scenario. More specifically, Openness and Agreeableness depend on the dynamic information, but only for some avatars. As can be seen in the graphs on top right of Figure 5.3, *Obadiah* and *Spike* were more open statically, but *Poppy* and *Prudence* were either less so or were the same. When judging the avatars solely by their appearance, the male avatars were perceived more agreeable and the females less. For Conscientiousness, there was an effect of scenario ($F(1, 18) = 4.627$, $p < 0.05$) but there was no interaction ($F(3, 54) = 0.934$, $p > 0.4$). With exception of *Obadiah*, who was rated higher in Conscientiousness for his image than in the full interaction mode, all avatars were rated nearly the same in both scenarios. In the case of Extroversion, there was an effect of experiment and an effect of interaction ($F's > 4.6$, $p's < 0.01$). *Poppy's* static appearance got rated a lot lower, but the full interaction mode had no real effect on the perceived Extroversion for the other avatars. Finally, for Neuroticism, there was neither an effect of scenario ($F(1, 18) = 0.236$, $p > 0.6$) nor was there an interaction ($F(3, 54) = 1.381$, $p > 0.2$). Meaning that all avatars were judged to be as neurotic in the static experiment as in the full interaction mode. Note, that the swapping of quadrants mentioned above for *Poppy* map to a statistically significant

change in the actual Extroversion rating. Moreover, although the other three avatars did not swap quadrants, the male ones did experience significant changes in their apparent personality.

In sum, the cross-experiment analysis shows that the *FFMRF can pick up subtle changes in personality* and the coarse quadrant-analysis approach masks visible effects. Moreover, the results show that although the static information certainly was used, *the actual behavior of the avatar clearly affected the perceived personality.*

5.6 Experiment 7: Effect of Physical Channels

Clearly, some of the information we use in order to assess the personalities of the ECAs relies on the dynamics or the interaction. In this experiment, we examined the relative importance of visual and auditory interactions. One group of participants had the "Telephone" scenario, where they could hear – and be heard by – the avatar but no visual information was transferred. The second group of participants had the "Glass Wall" scenario, where they could see – and be seen by – the avatar but no audio information was transferred.

Overall, *the scenarios affected the different characters in different ways.* When only visual feedback was available, only *Poppy* kept her desired quadrant in Eysenck's personality space. The sign for Neuroticism was opposite to the designers' intention for *Obadiah* and the sign for Extroversion was also putting *Prudence* and *Spike* on the undesired side. For the audio interaction mode *Obadiah* and *Poppy* got placed in the intended quadrants of Eysenck's space. *Prudence* and *Spike* had the right half for Neuroticism but their Extroversion sign was again incorrect (see Figure 5.2, comparing the blue and green symbols in Eynseck's space for a more accurate position). By examining the sub-traits, some of the SALs can be placed in their desired quadrants. In the case of *Obadiah* only Anxiety (N1) and Depression (N3) should have been considered for the visual scenario (see the corresponding graphs on bottom left of Figure 5.3). In both scenarios, to adjust the Extroversion ratings, in *Prudence*'s case just Excitement-Seeking (E5) would have been enough and for *Spike* Assertiveness (E3) and Excitement-Seeking (E5) (and also Activity (E4) for audio) (see bottom half of Figure 5.3).

A two-way ANOVA with avatar as a within-participant factor and physical channel as a between-participants factor showed a main effect for character for all dimensions ($F(3, 54) > 7.65$, $p < 0.001$), a main effect of physical channel only for Agreeableness ($F(1, 18) = 4.909$, $p < 0.04$; all others: $F's > 0.095$, $p's > 0.1$), and an interaction for every dimension ($F's > 2.99$, $p's < 0.04$). In other words, the perceived personality depended both on the specific character and the specific physical channel that was available.

The ratings for each OCEAN dimension for each avatar are shown in the corresponding graphs in the bottom half of Figure 5.3. Both male characters sounded less Open than what was derived from the visual channel. For females, the effect was reversed. Note that this is consistent with the observations inferred from the comparison of the full interaction mode against the static one (which had no audio). *Spike* was the only SAL perceived to be more Conscientious in the experiment with audio; all the rest were considered more undependable. Extroversion seemed to be perceived the same in both scenarios for *Prudence* and *Spike*. *Obadiah* sounded way more introverted than he acted, while for *Poppy* was exactly the opposite. Only *Prudence* was judged more Agreeable when heard. *Poppy* was rated less and the males way more hostile in the acoustic scenario. For Neuroticism small changes were perceived in the cases of *Prudence*, who was rated a bit higher in the audio mode, and *Spike* and *Poppy*, who got a bit lower rates. *Obadiah*, on the other hand, was considered to sound way more neurotic than what people judged just out of the visual information. In sum, *the final rating of each personality dimension can not be assessed by just analyzing one communication channel.* Care needs to be taken when designing an ECA that both its visual and acoustic behavior are considered.

5.7 Unimodal versus Multimodal Personalities

To examine the effect of having access to only a single physical channel (i.e., the unimodal scenarios in Experiment 7) versus having all information (i.e., the multimodal scenario in Experiment 5), we did a repeated-measures ANOVA on the Likert ratings from Experiments 5 and 7. As in the previous analyses, avatar was a within-participants factor and scenario (Multimodal, Telephone, and Glass Wall) was a between-participants factor. Unlike the previous analyses, we now explicitly add OCEAN as a within-participants factor. Overall, the results are consistent with the individual analyses

of the two experiments, although there are some apparent differences. There was a main effect of OCEAN ($F(4, 520) = 7.239, p < 0.001$), a main effect of scenario ($F(2, 520) = 3.824, p < 0.03$), and no main effect of avatar ($F(3, 520) = 1.065, p > 0.3$). The main effect of OCEAN dimension is modified by a significant interaction between OCEAN and avatar ($F(12, 520) = 3.568, p < 0.001$). The main effect of scenario is modified by an interaction between scenario and avatar ($F(6, 520) = 2.554, p < 0.02$). Most critically, there is a significant three way interaction ($F(24, 520) = 2.683, p < 0.001$). In summary, the FFMRF can be used to measure the personality of virtual humans (main effect of OCEAN), different virtual humans have different personalities (interaction of OCEAN with avatar), and the exact personality a virtual human is seen to have depends on which semiotic channel is available (three-way interaction). This three way interaction can be seen in the bottom half of Figure 5.3. From the wealth of details available in the graphics of that figure it is clear that the multimodal versions of the avatars are seen as having personalities that are clearly different than either of the unimodal personalities. In other words, *when designing an avatar, not only should the visual and the acoustic behavior be considered, but also the interaction or consistency between them* (see also the work from Vinayagamoorthy et al. [72]).

5.8 Discussion and Conclusions

Overall, it is clear that despite the fact that computers are not animate and have no feelings, intentions, or beliefs, people can and do treat virtual characters as though they have human-like personalities. Thus, standard psychological models – and measurements – of personality are just as appropriate for virtual as for real humans. It is also clear that a more complex model like OCEAN can pick up rather subtle changes in the perception of an agent's personality, especially if the sub-scales are used. Furthermore, it is clear that information about personality can be found in every considered physical channel. In addition to altering the perceived personality of an individual through geometry and texture, the apparent nature of a virtual character can be altered through his or her behavior. In short, the appropriate choice of facial expressions, answers to our questions, or even the timing of answers, all greatly affect how we think of our virtual dialog partner, and how willing we are – or are not – to deal with them. Perhaps more interestingly the same aspects of personality are in different physical channels for different people. Finally, and most critically, in many cases the perception of an agent's personality when we can

both see and hear it (and it us) is not always a weighted sum of its visual and acoustic personalities. That is, not only is personality multimodal, but any attempt to study or design personality by solely focusing on voice or face characteristics will most likely fail. Moreover, the results have shown that targeted changes in one aspect of a personality (e.g., hostility) often brings changes in other personality traits (e.g., competence).

To exemplify the findings, we discuss the ratings in detail, across all experiments, for two opposite avatars: the passive-negative *Obadiah* and the active-positive *Poppy*. For *Obadiah*, when looking at his Openness, we see a clear trend (ratings of 3.1, 3.8, 4.1 and 2.9 for Full, Static, Glass Wall and Telephone conditions, respectively). That is, his facial motion made him appear somewhat open (4.1 is just above neutral). The fact that when just looking at his face, he is less open suggests that the facial motions are what was driving his Openness. His auditory personality was very closed, as was his personality in the full interaction mode. In other words, Openness is driven by the voice for *Obadiah*. For Conscientiousness, his ratings were 3.3, 4.3, 4.0 and 3.6 for the Full, Static, Glass Wall and Telephone conditions, respectively. His face (and to some degree, his facial expressions) looks a bit conscientious but his voice reflects a less reliable personality. Note his Full rating is lower than either the Telephone and Glass Wall conditions, but is close enough to the Telephone condition to say that Conscientiousness is driven orally for him. The low value of Extroversion he was supposed to have, was achieved by his voice and his facial appearance. His facial motions, on the other hand, made him seem more extroverted (ratings of 2.4, 2.4, 3.8 and 1.9 for Full, Static, Glass Wall and Telephone). The full-modality personality matches his appearance, but seems to be driven by acoustic information. He looks very Agreeable, although his facial motion decreases this impression a bit (ratings of 4.3, 5.1, 4.7 and 3.8 for Full, Static, Glass Wall and Telephone, respectively). Nonetheless his voice makes him seem quite antagonistic, and this is reflected in the full mode. *Obadiah* looked and sounded Neurotic (a tendency that was followed in the full interaction mode), but his facial motion attenuated this judgment (ratings of 4.9, 4.6, 3.6 and 4.8 for Full, Static, Glass Wall and Telephone).

When analyzing *Poppy*, we found similar tendencies for Openness. The Full condition seems to be driven by the acoustics (5.1, 3.7, 4.3 and 5.0 for Full, Static, Glass Wall and Telephone). For Conscientiousness we find that she is very conscientious in face and facial motion, but much less so acoustically. The multimodal personality seems to be a weighted sum (4.0, 4.6, 4.7 and 3.4 for Full, Static, Glass Wall and Telephone). She sounded Extroverted but her static face looked introverted, and the facial motion placed

her somewhere between (5.6, 3.8, 4.5 and 5.7 for Full, Static, Glass Wall and Telephone). The overall judgment in the multimodal case was really close to her auditory personality. For Agreeableness (4.1, 4.1, 5.0 and 4.9 for Full, Static, Glass Wall and Telephone) the ratings suggested that the two unimodal conditions have some unexpected interactions: *Poppy*'s voice and facial motion make her a somewhat agreeable person but her facial structure suggests she was rather neutral. The overall judgment in the multimodal case was again lower than either of the two interactive unimodal conditions. Finally, her ratings in Neuroticism (2.8, 3.3, 2.8 and 2.5 for Full, Static, Glass Wall and Telephone) also show a multimodal value that is exactly the visual personality.

As can be seen in Table 5.2, we can derive rough guidelines to indicate where the efforts of anyone interested in providing an ECA with a personality might go. For designing negative personalities, it seems like a focus on appearance is important. For active personalities, the audio channel is the one we should focus on in case we want to make the character more active. In the case of emotionally balanced characters the final perceived personality is an average of all the channels (with the looks being in second plane) and the channels to focus on depend on the desired personality, i.e. it could be useful to invest most of the efforts in the audio interactivity if the goal is to generate a trustworthy, competent avatar.

Avatar	O	C	E	A	N
Obadiah	Audio	Audio	Audio	Combination	Audio
Poppy	Audio	Combination	Audio	Static	Visual
Spike	Unknow	Visual	Visual=Audio	Audio	Visual
Prudence	Audio	Visual	Visual	Audio	Visual

Table 5.2: Dominance of unimodal channels in the avatars' perceived multimodal personalities.

In general, personality dimensions were unimodally dominant, but there were some cases where the full-modality rating was not predictable as a weighted sum of the unimodal values. In other words, it might be possible to design a virtual personality using information solely in a single channel, but perception of real personalities as well as the design of some virtual personalities requires simultaneous attention to multiple informational channels. Of course a closer examination of the sub-scales can provide more detailed insights. Likewise, further experiments where specific elements of the

different channels (e.g., head motion, eye motion, eyebrow motion, and mouth motion for the visual channel) are independently manipulated can provide much more exacting mappings between personality and behavior.

In order to be able to make more detailed conclusions about subtle personality aspects, about precisely which behaviors affect those dimensions and about the different influence of each channel on the perceived personality, the more detailed NEO-PI-R could be used, along with systematic changes in the ECA's behavior or appearance. Thus, ECAs can be a useful tool in examining the mapping between behavior and personality. Of course, such a mapping will also be useful in creating new ECAs for specific applications. It might soon be reasonable to have OCEAN personality profiles explicitly involved in the creation of ECAs.

PERSONALITY IS IN THE MOVEMENT: MAPPING SEMANTIC AND PERSONALITY SPACES

6

The use of videos in the experiments conducted in Part II allowed us to have different versions of the same emotion. Since different people have different personalities, they will perform the same expression differently. Thus, as already indicated in Chapter 2, an examination of how a given person's expressions differ from the mean can provide insights into their personality profile. Following this inspiration, and, given that our database contains a very large number of expressions, it was possible to perform an experiment were people measured the perceived personality from the actor by filling an standard, validated questionnaire. Thus, in this chapter, we use additional expressions, for a total of 62 on ten different people for whom we also gathered a measure of their personalities (or perceived personality traits). This allowed us to develop an initial mapping between the personality space and the Semantic Space for facial expressions, completing the pipeline from personality space through semantic expression space to actual behaviour and facial expressions, enabling style motion transfer.

This work is yet to be published and was done in cooperation with Philipp Hahn (Graphic Systems Department, BTU Cottbus-Senftenberg), who helped conducting the experiments and the advisor of this thesis, Prof. Dr. Douglas W. Cunningham (Graphic Systems Department, BTU Cottbus-Senftenberg), who give advise concerning the analysis of the data.

6.1 Introduction

To develop an ECA which enables a meaningful and successful communication with the user, some design decisions must be made first. Among other factors, an appropriate embodiment of the ECA is important and must be chosen wisely according to its function and abilities. The software agent can be for example a living creature, an animal, an inanimate object or a human being. The level of realism regarding its appearance is as well an important aspect, would an cartoon-like ECA be enough or do we want and need perfect realistic hair, eyes and skin [89]. Mental abilities also need to be considered; we suggest that a successful and meaningful communication is defined by exactly two factors, the domain-specific knowledge, which entitles the competence of a system in relation to its function and the social intelligence, which describes the ability to get along with others. Skills as listening, empathizing and expressing emotions are important for a realistic user interaction, thereby, we need to decide in which sense such social skills help the user or distract them, and, if designing the ECA to be emotionally driven by its personality would be a good option.

Once the computer is granted a virtual body, it must be given the ability to use it to non-verbally convey socio-emotional information (such as emotions, intentions, mental state, and expectations) or it will likely be misunderstood. Despite the fact that machines do not process nor produce socio-emotional information, people still tend to use interpersonal behavior cues when interacting with them [49, 50]. It has been demonstrated that people respond to these cues analogously to the way they will do to another person [50, 72, 90, 91, 92]. If we are striving for a human machine interaction as realistic as possible, we want to consider mental abilities for an ECA as they could help to improve the communication drastically. Among them, personality seems to be almost indispensable: it reinforces the user, strengthens the bound between user and computer, increases the tolerance for mistakes and overall makes the communication more efficient and effective because the ECA's actions, intentions and wishes can be derived from it [72]. This is based on the fact that non-verbal communication seems to be key when trying to form a opinion on the true psychological sate of an individual [93], as it is one of the main channels to project personality in an unconscious way [94, 95].

As we showed in the previous chapter, social psychology research on the assessment of personality can indeed be applied to virtual agents. Our results indicated that the

apparent nature of a virtual character can be altered through his or her behavior. Thus, the appropriate choice of facial expressions, answers to our questions, or even the timing of answers, all greatly affect how we think of our virtual dialog partner, and how willing we are – or are not – to deal with them. Moreover, the results showed that targeted changes in one aspect of a personality (e.g., hostility) often brings changes in other personality traits (e.g., competence). Thus, it is clear that modeling a personality is not trivial. But, what if we would be able to directly transfer the personality of an individual to an ECA? The Motion capture technique does capture the specific movements done by an individual, thus, because of all the aforementioned reasons, it also captures those variations from the generic expressions that help conveying the personality of the actor. It is clear that all channels are important to be able to convey the desired personality, as we showed in Chapter 5. That is, personality is multimodal. Thereby, even if while designing the ECA we need to be aware that different appearance and voice characteristics will affect the impression on the personality given to the interlocutor, and we need to be careful to make them match the ECA's intended personality, having a mapping that establish the relation between facial movements and OCEAN traits can be really helpful for the design of the virtual agent. This mapping would simplify the work of the designers, as they would be able to both generate the desired expressions for the ECA, and provide it with a chosen personality while only needing to deal with one set of parameters, as the differences required to model the personality will reflect on simple swifts on the desired location of the expression in the Semantic Space.

Nevertheless, the mere existence of this mapping is based on the assumption of the existence of such a relation between personality and facial movements. Previous chapters' results pointed towards this direction, allowing us to hypothesize that the perceived differences among different people expressions were related to their personalities. To test this hypothesis, this chapter needs to address first the following research questions:

- Are there cultural dependencies in the perception of meaning, when analyzing the expressions of German and Spanish individuals?
- Are the differences perceived among the expressions of different individuals a reflection of their personalities?

This chapter extends the work presented in the previous chapters to solve these questions and set the basis to find the desired mapping between the Semantic Space for facial expressions and the Personality space.

6.2 General Methods

In the following we describe the experimental procedures used in this chapter. Note that they are all consistent with the procedures used in all previous experiments presented in this thesis.

6.2.1 Stimuli

In order to extended the Semantic Space for the found in Chapter 3, Section 3.3.2 and being able to determine the averaged position of the full range of expressions recorded in our DB as if performed by a generic actor, we used as stimuli for the experiment presented in this chapter the full collection of real videos (please refer to Table 3.1) corresponding to our ten recorded subjects (five male, five female) (see Figure 3.1).

6.2.2 Psychophysical Methodology

Scales

Naturally, to keep the consistency with the rest of the experiments presented in this thesis, we used the same scales and methodology proposed in Part II. We kindly refer the reader to Table 3.2 for a reminder of these scales and their corresponding factors.

Personality Questionnaire

The results of Chapter 5 confirmed the validity of the FFMRF to measure both human and virtual personalities. Coherently with this last chapter, we also chose the OCEAN model to describe personality and used the FFMRF in the following experiment (see Appendix B). The detailed ratings of a personality along the sub-scales described in Table 5.1 this personality questionnaire offers, help to better measure and study the subtle changes in the perception of personality.

PROCEDURE AND DESIGN

We performed a single experiment, with two different tasks, to both recover the SSp underlying the full spectrum of emotions contained in our video database and obtain the perceived personalities for all the recorded individuals.

Even when in Chapter 3 we argued and proved that it was unnecessary to validate the full database, as one actor proved to be enough to demonstrate the stability of the video SSp, this chapter's goal justified the investment of the resources. In order to find the mapping between personality and the SSp, we needed to analyze the deviations from the center for more than one person and we decided to take this opportunity to fully evaluate the database for all actors and all expressions.

As it was already mentioned in Chapter 3, for the semantic differential task to produce reliable results, the number of trials per experiment should not be too large (for an ideal of 600 trials [21]). Evaluating an actor requires a total of 744 trials (62 expressions along 12 scales), and consequently a total of 7440 trials to evaluate the full database. Therefore, and willing to keep a full within-participants design, we divided our experiment into 10 sessions equally spaced along 10 consecutive days for each participant.

A total of 10 people (5 females, 5 males, age range $20-31$) participated in our experiment and were compensated with 8€ per hour for their participation. The average time a participant took to complete one session was 1 h and 18 minutes.

Before being asked to fulfill an informed consent form and without revealing the research question behind the experiment, each participant was instructed on how the experiment would run and was giving the chance to ask any questions. They were informed that they could stop the experiment at any point without any negative consequences to them.

For each session, each participant was seated in a properly isolated space of a semi-dark room (to provide proper seclusion from other participants that could be in the same room), roughly 50 cm in front of a 24" LED monitor (at a resolution of $1920x1080$). For the first session for each participant, they were presented with a screen with the instructions for the experiment and, after having asked a control question to ensure that the participant understood the tasks of the experiment, the experimenter left the room.

Similarly to the experiments conducted in Part II, the experiment was controlled by Psychophysics Toolbox Version 3.0.11 (PTB-3) [40, 41, 42]. Once more, as all participants were native Germans, all stimuli and instructions were given in German.

The experiment consisted of two tasks, the semantic differential task and the FFMRF. During a session, the participant was asked to complete both tasks for the videos corresponding to a randomly assigned actor, with each participant receiving a different random order. For each trial of the first task, the procedure was identical to that explained for Experiment 3 in Section 3.2.2. Under the always visible main question: "How likely is it that these emotional features also occurred?", the left of the screen displayed a Likert scale from 1 to 7 anchored by a pair of terms from Table 3.2 while the right side showed a video for one expression of the actor assigned to the current session of the experiment. A snapshot of this interface is shown in Figure 3.3. By clicking on the corresponding number reflecting their desired answer the participant could move to the next trial. The order of appearance of the videos was randomized for each participant and session, but once a video was selected, it needed to be rated among all 12 scales before the next video was displayed. As second task of the experiment, after the participant had rated all the videos corresponding to the session, this is, the 62 expressions for an actor, the participants were asked to fill out the FFMRF for the actor.

6.3 Experiment 8.1: Real Videos

As previously stated, the main purpose of this chapter is to analyze the correlation between personality and Semantic Spaces. Towards that goal, the first step was to gather the data that would allow us to study how the differences between individuals while using facial expressions relates to their personality traits and if we would be able to replicate this personality while generating new facial expressions by just shifting the desired location in the SSp of the desired expression according to the targeted OCEAN profile. In order to do so, we followed the methodology described in Section 6.2. We recovered the SSp defined by the 62 facial expressions (in their video form) for our ten actors presented in Chapter 3 using the same methodology proposed in Chapters 2 and Chapter 3 and the personality questionnaire used in Chapter 5. After recovering the SSp and analyzing the deviation of each individual actor to the average center of expressions,

we studied the correlation between this deviations and the gathered OCEAN profile for the individual.

6.3.1 Recovering the Semantic Space

Analogously to the analysis conducted in the experiments of Part II, we consulted a number of different methods to decide on the number of necessary factors to explain the variance in the data. Theoretical reasons and the Explained Variance Criterion (for almost a 90% explained) would argue for a four-factor solution. The scree test criterion suggest a three-dimensional solution while the parallel analysis, optimal coordinates, acceleration factor and Kaiser criteria agreed that two factors would suffice.

Exploring all the options through factor analysis, the amount of variance explained does not only depend on the number of factors considered but also on the rotation applied. While promax rotation is able to explain a 75.3% 77.2% and 83.3% of the variance for the corresponding 2D, 3D and 4D solutions, applying either no rotation or varimax rotation increases these percentages to 75% 80.6% and 87.2% respectively. A closer examination of the loadings for the 3D and 4D solutions makes clear that the increase of considered dimensionality is not justified and leads to incoherent groupings of the factors. Both 3D and 4D solutions consistently recover the Valence dimension by fusing the scales corresponding to Evaluation and Potency, and generate the Activity dimension. Then, 3D solutions fuse this last factor with the Predictability factor to create the Arousal dimension, leaving one dimension without proper loadings, while 4D solutions tend to split the scales corresponding to predictability on both the remaining unloaded factors. Furthermore, when forcing 3D and 4D solutions while applying promax rotation, Heywood cases occur, reinforcing the theory that we are trying to extract too many factors.

The 2D solution (which explains in between 75% and 75.3%of the variance depending on the applied rotation) recovers the same fusion of the EPAP dimensions we saw in our previous experiments. Once more, the exception on the correct loading onto the expected dimension is scale 8, which formulation we decided not to alter to keep the consistency between experiments. Both varimax and promax rotation perfectly load scales 1, 5, 9 (Factor Evaluation) and 2, 6 and 10 (Factor Potency) onto the Valence dimension while all the rest of the scales (except 8, as already mention) are loaded into the fusion of

Predictability and Activity, thus, forming Arousal. The variance explained by each of the dimensions for both rotations is quite similar: Valence explains 47.6% and 47.8% for varimax and promax respectively, while Arousal explains 27.4% and 27.5%. Using no rotation evens the amount of variance explained by each factor (37.2% and 37.8%) but loads scale 10 into Activity instead of making it fall into Arousal as desired. Table 6.1 shows the factor loadings gathered through factor analysis with promax rotation, giving its ability to successfully recover the desired dimensions and its marginally superior performance in comparison to the use of varimax. The recovered Semantic Space is shown in Figure 6.1.

Scale ID	**Factor 1**	**Factor 2**
1	**0.968**	-0.039
2	**0.917**	-0.040
3	-0.527	**0.837**
4	-0.264	**0.718**
5	**0.962**	-0.118
6	**0.809**	-0.119
7	0.316	**0.905**
8	*0.199*	0.149
9	**0.942**	-0.088
10	**0.888**	0.144
11	0.462	**0.829**
12	-0.051	**0.706**

Table 6.1: Factor loadings for the extended 2D space for the 62 expressions recovered through promax rotation. The numbers in bold show the significant contributions.

Despite the complexity of Figure 6.1, a number of things are clear after careful examination. The recovered Semantic Space shows a nicely clustered distribution of the 62 expressions. In general, similar expressions are closely nested, as is the case to most of the types of smiles and thinkings. Positive emotions are placed on the right of the space, indicating their positive valence, while those expressions with negative attributes are consequently placed on the left. Sudden, ephemeral, unexpected expressions are located on the top hemisphere of the space, while slower, more permanent expressions are located on the lower half of the space. Moreover, the relative location between different variations of the same expression are located as one could expect. For example, the location of "Surprise" (Surp) defines it as rather unpredictable and fast, but neutral

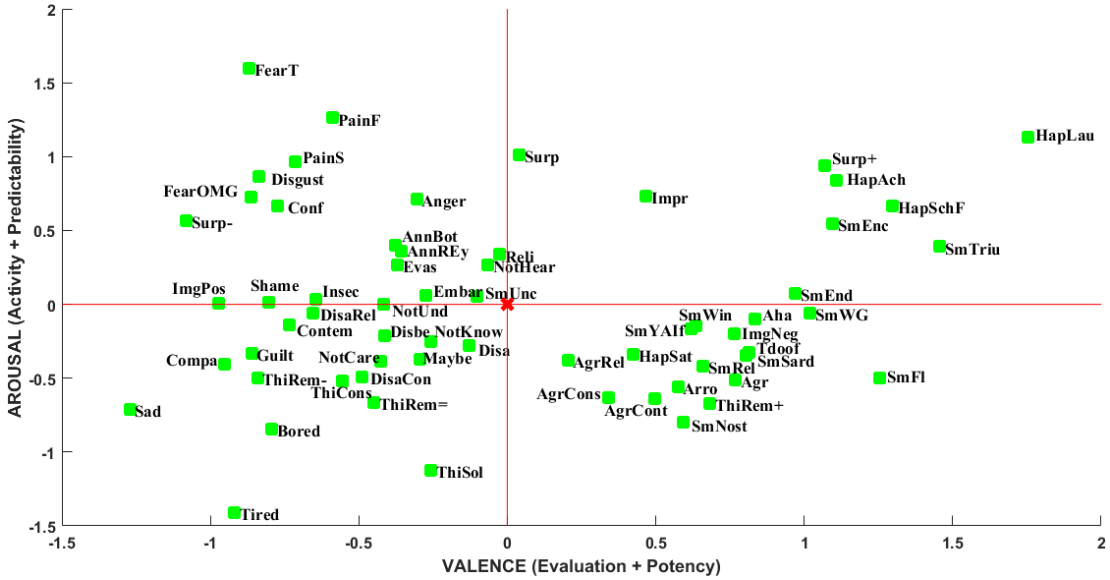

Figure 6.1: Coordinates along Valence (Factor 1) and Arousal (Factor 2) for the averaged positions of each of the 62 expressions among actors.

in valence. The positive and negative versions of this expression – "Pleasant Surprise" (Surp+) and "Unpleasant Surprise" (Surp-) –, are coherently placed on the right and left sides of the neutral version while preserving the arousal value.

6.3.2 Comparison with Previous Experiments

When analyzing the individual differences between individuals for a given expression, it was clear that every actor performed every expression in a unique way. This is illustrated in Figure 6.2 and is coherent with the findings of Experiment 2 (Section 2.4). These differences were naturally bigger for some expressions, such as "Tired" (Tired), while some others were quite similar among our actors, such as "Sadness" (Sad). The variation among individuals for "Unpleasant Surprise" (Surp-) seemed to be specially big. Further from seeing this as a negative point, we found this large variance to be a clear example of our theory on how our personalities color our emotions. The particular scenario used to trigger this emotion on the individuals was one where they would realize that they have lost their wallet (see Appendix A). During each individual recording session, this scenario proved to be effective to trigger the desired expression on every actor. Nevertheless, when visualizing the recordings for all individuals, there were clear differences showing

that the relevance of the consequences (or the consequences themselves) such a scenario would pose, were quite different for each person. For example, one could think about the money loss or even about what would it imply, like not being able to have spare money to spend on entertainment, or not being able to survive the month. Also, one could be more worried or annoyed by the lost of important personal documentation or cards, with the subsequent risk and/or the annoyance of needing to cancel and renew them. Another example could be the variance found on "Happy Achievement" (HapAch). The level of happiness reached by finishing a task and achieving a goal could very much depend on the personal level of rigorousness, as for a person who pursues everything to the point of perfection achieving it could mean a lot or. Also, the intensity and nuances of this expression could be determined by the self-esteem of the individual, boasting it with a feel of pride, or diminishing it due to the lack of confidence.

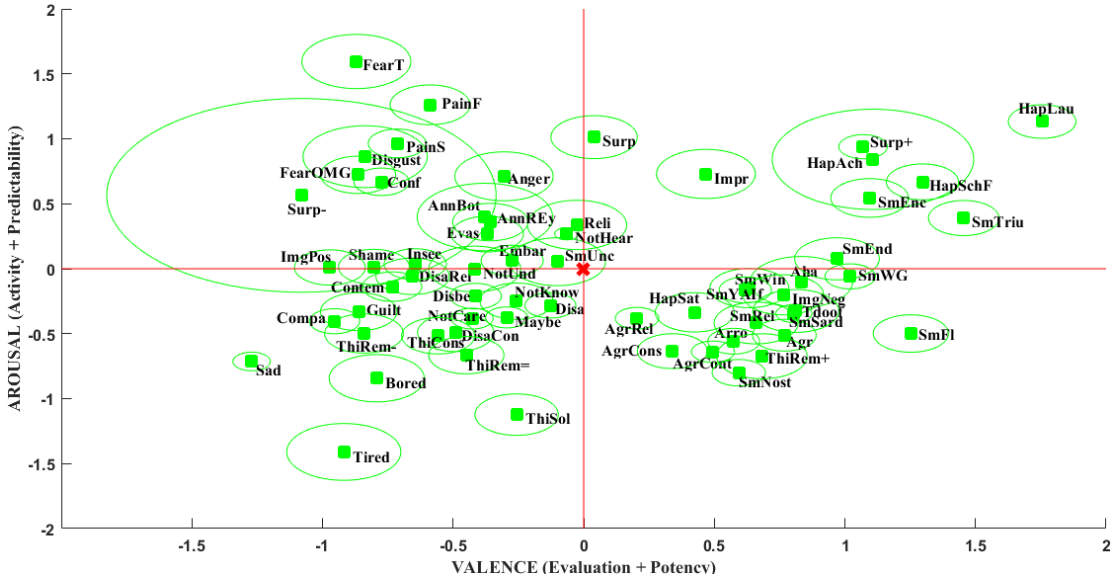

Figure 6.2: Coordinates along Valence (Factor 1) and Arousal (Factor 2) for the 62 expressions averaged among actors. The squares represent the video results and the radius of the ellipse around each emotion is the SEM for euclidean distance of each actor to the mean expression.

In Chapter 2 we argued that the variance in the way to express a given emotion or state between the different German recorded individuals from the MPI facial expression database [10] was related to their personality. The results from Chapter 3 reinforced this theory, as analyzing the expressions for one Spanish actor from our recorded database and projecting them into the original SSp derived from the German actors, did not

alter the mean Euclidean distances to the averaged positions of the expressions. This discarded the possibility of the differences showed for one individual to the average expression to be due to the information present in the videos (e.g. different background or illumination) or, most critically, to cultural differences (at least when comparing these two European cultures). Even when it would be valuable to extend this study by including new databases of a broader spectrum of people (e.g. Asian, American or African individuals) to analyze intercultural differences, we can conclude that for our data, the variance was merely due to personality traits.

To further prove this, we directly compared the results of our new experiment to those in Experiment 2 (Section 2.4). In order to compare the location of common expressions of both experiments and DBs in the defined SSps, we first averaged each expression across all actors in order to find the averaged positions of all expressions on each SSp, the one defined by the 9 expressions in the MPI DB and the one we just presented containing 62 expressions. Next, we derived the spaces underlying these two sets of data, using factor analysis in the case of our most recent experiment and PCA on the case of the MPI videos. All criteria suggested a 2D solution for both spaces. After having defined them and selected from the 62 expressions space the common 9 expressions with the MPI space, we proceeded to examine their correlation. A procrustes analysis gave a distance between the two matrices of $d = .1946$, yielding a significant correlation of $r = .8974$ ($p < 0.001$) between them. Thus, *we can confirm that most of the variance is due to personality.* Very minor differences could be due to cultural differences and we think it would be worth to conduct a more detailed study to examine them. One could argue that the absence of cultural differences could be (among similarly enough European cultures) due to the culture the participants belong to and not to the one of the recorded individuals.

6.4 Experiment 8.2: OCEAN Profiles

As previously described, we recovered the OCEAN profiles for each of the actors of our database by asking the participants of our experiment to fulfill a FFMRF questionnaire for each individual after having observed and rated all the corresponding 62 videos for his or her recorded expressions. The gathered OCEAN profiles can be seen in Figure 6.3 and their detailed version in sub-scales are available in Figure 6.4.

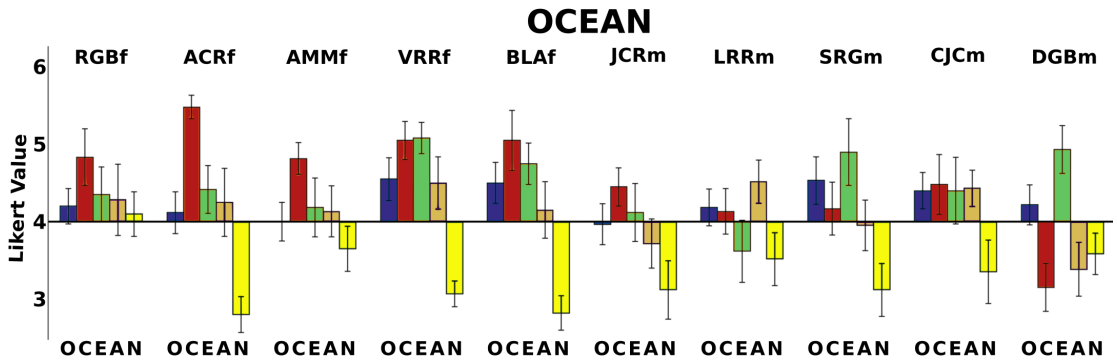

Figure 6.3: Averaged ratings for the Big Five Factors for each actor.

The figures illustrate how each actor was indeed perceived to have a different personality. There were also some interesting group trends, like for example, gender ones. We can see how all female individuals (identifiers ended by "f", left part of the plots) had a tendency to be perceived as more practical than the male ones (identifiers ended by "m", right part of the plots) and also were perceived as more Conscientious than men, with the extreme case of "DGBm" who was the most undependable among the male actors. This individual also did not follow the general common trend of all the rest of the individuals, being rated as more antagonistic than the rest. As was also discussed in Chapter 5, some very interesting details given by the sub-scales were masked by the average rating for each OCEAN scale when defining an individual. For example, when observing the Neuroticism averaged rate of "RGBf", she was considered almost as neutral (averaged score = 4.1), nevertheless she was perceived as anxious (N1 - Anxiety) and rather optimistic. "DGBm" scored in average for this dimension as slightly under the neutral (3.6) but the sub-scales revealed that he was perceived as a shameless and impulsive individual (low N4 - Self-Consciousness and high N5 - Impulsiveness). Even a more clear example of this masking effect of the average is the case of "JCRm" and "SRGm". Even when both individual had an identical averaged score of 3.1, the sub-scales revealed that they were not perceived the same. The latter was perceived to be a more impulsive person (N5 - Impulsiveness) while the former strike more as a timid (N4 - Self-Conciousness) and angrier character (N2 - Angry Hostility). Similarly, the average rating for "AMMf" in Openness was a perfect neutral (averaged score = 4) while a look at the sub-scales would reveal a total different story. This person scored quite low on O1 - Fantasy and quite high on O3 - Feelings, being perceived as a practical, self-aware person.

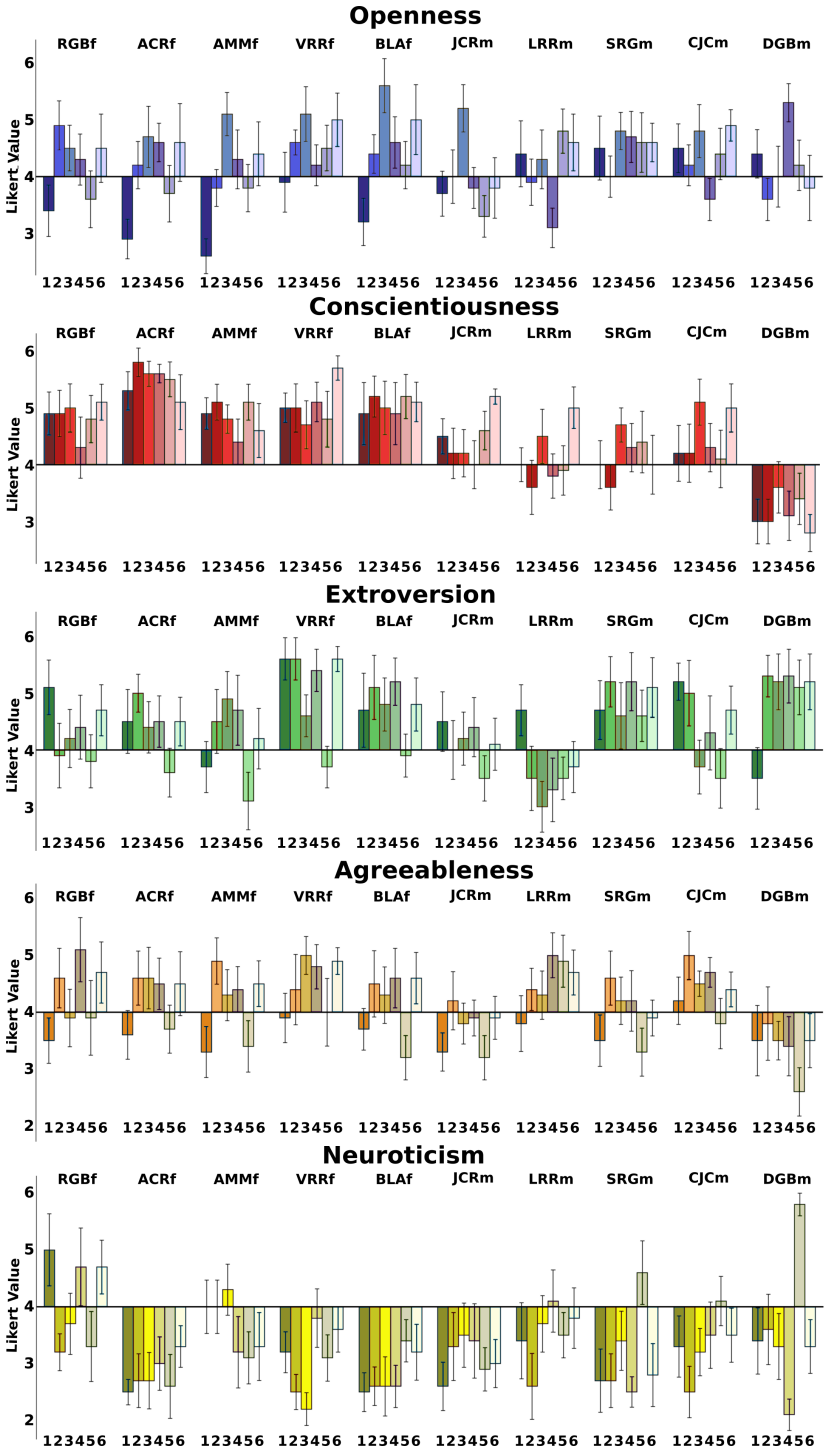

Figure 6.4: Scores for all 30 scales for each actor (error bars represent the SEM).

In this chapter we will focus solely in the general OCEAN ratings (see Figure 6.3), partially due to the sample size of our stimuli but mostly because that is the way the OCEAN model was designed to be used. Nevertheless, it would be worthwhile to consider a larger scale study with more individuals to be able to conduct a more detailed analysis on the trends of the sub-scales and their relation to facial expressions. That is, the relatively low number of participants means that the variation in ratings on the sub-scales will be too large for specific significance differences to be interpreted with any degree of confidence. Since it is very much worthwhile examining the variations on the sub-scales, running more participants should be considered in future work, potentially with the will 240 question version of the OCEAN questionnaire.

6.5 Correlation between Semantic and Personality Spaces

In order to achieve our goal of finding the mapping between the personality space and the SSp for facial expressions, we are interested in analyzing the deviation from the average position of the expressions that each recorded individual of our database showed. To help better visualize this, in Figure 6.5 we show the centers of the emotions among each actor in comparison to the main center of all averaged emotions. We used the Pearson Correlation Test to analyze the correlation between the location of the individuals' expressions in the SSp and their OCEAN profiles.

Note that, as a given location (x,y) in the SSp actually reflects its values for Valence (x coordinate) and Arousal (y coordinate), we will treat these two coordinates separately, in order to establish the mapping between the two dimensions of the SSp and the five dimensions of the OCEAN model.

A first analysis comparing the averaged center of expressions for each actor (see Figure 6.5) with their ratings in each OCEAN dimension showed that Conscientiousness clearly had an influence on the way people express emotions, there was a significant correlation between this OCEAN dimension and Valence ($cor = 0.6801273$, $p - value = 0.03046$) and a negative correlation with Arousal ($cor = -0.6983821$, $p - value = 0.02468$), indicating that the more Conscientious an individual was perceived to be, the more strong and positive were their expressions and the less sudden they appear, holding

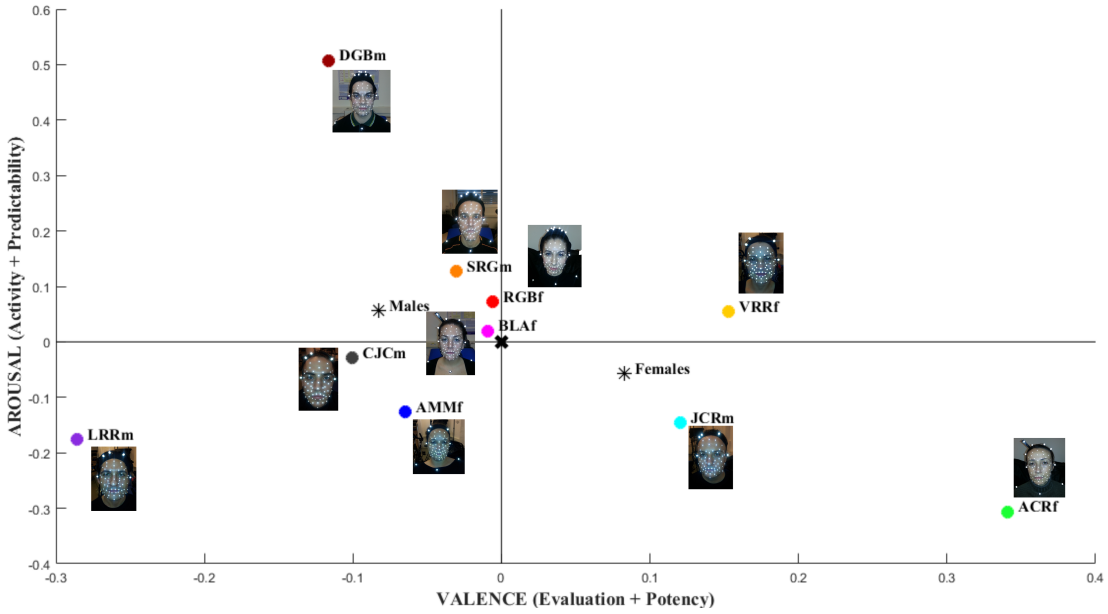

Figure 6.5: Centroid of emotions for all actors vs the individual centroids of each actor and the ones grouped by gender.

them for a longer time. Extroversion also showed a positive correlation with Arousal ($cor = 0.630615$, $p - value = 0.05061$), indicating that extroverted individuals were more prone to show fast emotions and faster to switch between mental states. Finally, there was a marginal negative correlation between Agreeableness and Arousal ($cor = -0.5774$, $p - value = 0.08048$) which could point out that highly Agreeable people would hold longer an expression and would be less predisposed to show fast emotions, perhaps trying to figure out the mental state or intentions of their interlocutor before expressing themselves.

It is worth to mention that, as what was compared in this analysis was the averaged position of an actor (i.e. the average across expressions within an actor) with the "grand mean" of emotions (i.e. the average across actors and expressions), it could be argued that what we were judging the "neutral" expression of a given actor against the generic neutral expression and, thus, this previous analysis could have given us the relation of the looks of a person with their evaluated personality based on dynamical stimuli. This would mean that, indeed, the appearance of a person influenced their perceived personality, as was also found in Chapter 5. This implies that, when designing a new ECA according to a desired personality profile, we should make sure that is their appearance makes them

look more contentious, extroverted or even agreeable, their facial animation should be adjusted accordingly (and vice-versa).

In the interest of finding a correlation between the OCEAN model and the actual dimensions underlying expressive facial motions, we also performed the analogous Pearson Correlation Tests between Valence and Arousal and the OCEAN dimensions considering the individual locations of all emotions for all actors with their ratings on each personality dimension. The analysis found that all OCEAN dimensions significantly correlated with Arousal. On the one hand, Openness ($cor = 0.0938$, $p - value = 0.0194$), Extroversion ($cor = 0.176$, $p - value < 0.001$) and Neuroticism ($cor = 0.08659959$, $p - value = 0.03108$) correlated in a positive way showing that a higher ranking on these personality dimensions would lead to a more impulsive form of facial expressibility. On the other hand, Conscientiousness ($cor = -0.1955$, $p - value < 0.001$) and Agreeableness ($cor = -0.1616$, $p - value < 0.001$) correlated negatively with the fusion of Activity and Predictability, confirming the tendencies observed in the aforementioned analysis. Finally, three of the OCEAN dimensions showed a statistically significant correlation with the fusion of Evaluation and Potency. Conscientiousness once again was positively correlated to the strength and positivity of expressions ($cor = 0.18049$, $p - value < 0.001$), as was Extroversion ($cor = 0.0889$, $p - value = 0.02685$) meaning that extroverted people could tend to show a more positive version of the emotional state they experience and also reflect it in a stronger way. Last but not least, Neuroticism had a negative correlation with Valence ($cor = -0.1434$, $p - value < 0.001$), thus a low score in this OCEAN dimension would have similar impact on the way of showing facial emotions as a high rating on Extroversion.

In other words, when conversing with a methodical, ambitious, reliable, efficient, devoted and/or reflective person, it is just but to be expected that they would tend to be more assertive, having conviction on their reactions. It would be expected from conscientious individuals a tendency to reflect before acting, thus being less prone to sudden expressions.

The higher a person rates on Neuroticism, the shorter, more sudden, weaker and more negative will be the expressions this individual will show. This seems to be reasonable, as an averaged high raking on this dimension will indicate a low emotional stability with clear negative traits. It is clear the expressions of an angry and/or depressive person will tend to be negative and probably volatile. Also, those from a timid and fearful individual,

will lack the strength and durableness that a self-assured person could convey. We would expect to see the same tendencies on introverted or undependable individuals regarding weakness and negativity, even when the reasons behind this behavioral tendencies could be quite different. In contrast with neurotic individuals, introverted ones would show longer and less abrupt expressions, as one could expect from a quiet, indifferent, lethargic, cautious and/or placid person.

It could be argued that the level of Agreeableness of a person would determine their tendency to mimic other persons' reactions or to adjust to what they think people expect from them. Following this same line of argumentation, the level of Openness to one's experience would follow the same trend with the only difference of having one's own expectations as a guide. Then, these two personality dimensions should not alter the valence of the individual's expressions in order to match the corresponding model. On the other hand, we also could expect that this difference on setting the standard to follow will affect the arousal of the expressions. An agreeable person should observe and reflect before reacting, decreasing the rapidness of their expressions. Also, either in order to be sure that their reactions are perceived, or because of the high emotional load to process not only one's own feelings but also those of the interlocutor, the expressions could be longer held as switching mental states could take more time. On the contrary, Openness would have the opposite effect on Arousal. We could expect this lower arousal on the expressions of a closed person. The more pragmatical, quiet, passive, rigid, and/or inflexible a person appears to be should be reflected on meditated, calmed, and stable expressions.

6.6 Conclusions and Future Work

In this chapter we have extended the work presented in the previous chapters, by expanding the Semantic Space (to define all 62 expressions for all 10 individuals from our database), by providing a personality profile for each individual, and by making a first attempt to map the Semantic and Personality spaces.

From our results, we can confirm that there were no cultural dependencies in the perception of meaning, when analyzing the expressions of German and Spanish individuals. Moreover, our results also indicate that the differences perceived among the expressions of different individuals were, indeed, a reflection of their personalities. Thus, we can

conclude that it is possible to find a mapping between Personality and Semantic Spaces when only considering facial movements. Of course, such mapping will only refer to the perceived personality through visual information. To capture the full conveyed personality of a complete ECA that would be able to speak, this mapping should be completed with the corresponding study for audio communication and the interaction of both communication modalities, as we mentioned in Chapter 5.

In order to provide an ECA with a personality, we should be able to break down how much of each of the perceived traits comes from which channel. The main problem we face then is that this requires detailed experiments with a systematic variation of only one parameter at a time, studying the influence it has in the overall perceived personality. This can not be performed by real humans, due to the fact that one person has a set of characteristics which are intrinsic to themselves and not modifiable (looks, gender, age) and, of course their personality itself. ECAs, on the other hand, are the perfect tool for the emulation, as they allow us to systematically vary all the aspects in a fully controlled environment. Performing personality analysis for those virtual agents would be a great help for their design. This way, the ECAs could more easily reach the goal (or fulfill the intention) they were designed for. As shown in Chapter 4, the Semantic Space can be used to generate new facial expressions. Finding the desired mapping between this space and the Personality Space, we should be able to generate expressions that convey the desired personality profile, providing all needed tools for the aforementioned emulation.

The gathered results in this chapter have great potential for further analysis, which we plan to exploit in the future. One could study, for example, trends on age and gender, both on pure facial movements, personality and/or on the combination of both. Also, only regarding the Semantic Space, having now data for more than one actor would allow us to make further conclusions about which facial areas' movements correlate best with which dimensions. One could, e.g. use PCA on the markers' trajectories to see which markers are fundamental for which expressions. Also, using the data for quadrant- or cluster-based analysis could further improve the characterization of the spatio-temporal structure of the facial expressions, i.e. the characterization of the Semantic Space's dimensions. On top of that, given the measured personality profiles for our actors, performing a more detailed analysis of the sub-scales would be interesting. One possibility will be conducting a PCA on the gathered scores on the 30 personality traits for all 10 actors to analyze the nature of the underlying space. It will be worth

to test if such space is still 5-dimensional, and if the sub-scales still load on the known OCEAN dimensions. The recovered Personality Space could be different, given that the stimuli use for the FFMRFs were purely visual (people showing facial expressions, without interaction with the participant or audio information). There is clearly much more that can be done with the gathered 4D trajectories $\times$ 72 markers $\times$ 62 expressions $\times$ 10 actors $\times$ 30 personality traits.

Part V

Conclusion

CONCLUSIONS & FUTURE WORK

7

This thesis aimed to contribute to the design of virtual characters by offering a methodological approach to improve their facial animations and capability to convey a personality. Towards this goal, we created the Semantic Space for conversational facial expressions, studied its mapping to actual facial motions and examined the correlation between the Semantic and the Personality Spaces.

In the following, we provide a summary of the conclusions for each of the main parts of this thesis as well as a brief overview on current projects derived from this thesis and exciting new options for future research.

Facial Expressions In Chapter 2 we ran a standard semantic differential task, using scales and emotion words derived from Fontaine et al. [18], along with six new conversational expression words. We successfully recovered the same four dimensional (4D) space found by Fontaine et al., validating this way our changes on the methodology to recover the SSps. We then used the same task with video sequences of nine conversational expressions, each recorded from six people. Factor analysis found that two dimensions were sufficient to describe the variance. We also found that space for words and expressions were very similar, confirming that the perceived expressions in the videos were correctly labeled and, thus, allowing us to interchangeably use videos or words. We dedicated Chapter 3 to extended this Semantic Space to contain more facial expressions (up to a total of 62). The high correlation of our results along experiments confirmed not only that the new expressions were correctly recognized but the found Semantic Space

is easily expandable using between participant designs, as it was proof the easiness of projecting any new recordings into the existing space. In this chapter, we also empirically derive the mapping between facial motions and the founded Semantic Space by making use of MoCap data, as it seemed the most natural way to obtain a spatio-temporal description of the facial movements contained in the videos. We found a significant correlation between the location of the videos of the expressions to the MoCap recordings, that could be improved by augmenting the MoCap data with additional modalities such as eye tracking to more accurately reflect human socio-emotional behavior.

MOTION SYNTHESIS Chapter 4 proposes a simple technique to show the generative capabilities of the Semantic Space even in their current form. This technique combines, on a frame-by-frame basis, the Motion Capture recordings obtained in Part II to synthesize a novel facial expression. To provide such expression with the desired emotional tone, we use its location on the Semantic Space to find the distance to all other expressions in that space and use them as weights to combine the corresponding Motion Capture recordings. This technique, together with the findings on the following part, offers a promising research platform for the study of facial communication in a controlled, systematic fashion.

PERSONALITY Notwithstanding that part of providing ECAs with full human-like communicative capabilities is to give them a personality, and that the only aspect on their design that this thesis is addressing are facial movements, there were some open questions that we wanted to empirically prove before. Thus, in Chapter 5 we confirmed that people do treat ECAs as though they have human-like personalities that can be modeled and measured using standard psychological models, such as OCEAN. Also, we found that even when personality is multimodal, every physical channel does provide information about it and, therefore, is worthwhile to find the mapping between visual behavior and personality. Towards that goal, in Chapter 6 we extended the Semantic Space for the video expressions with the data from all ten actors recorded in our database, for whom we also gathered their OCEAN personality profiles. The analysis of the data and the comparison with the results from previous chapters confirmed that the deviations to the generic expressions shown by each actor where, indeed, related to their personalities and not to cultural differences. Finally, considering each actor's deviation from the

average while performing all and each expression, we found the correlations between OCEAN and our Semantic Space.

FUTURE WORK Aside from the possible improvements for each step of the corpus of this thesis – already mentioned in the corresponding chapters – here, we briefly describe some possible future work directions, some of which are currently work in progress.

A core part of the thesis was recording, cleaning and parsing the MoCap DB. During the data post-process, we realized that existing cleaning techniques were not suitable for our requirements, as they removed signal as well as noise, and new methods are still needed. Our lab has begun to develop a multi-scale, cognition-inspired, cleaning spatiotemporal MoCap cleaning algorithm in collaboration with Dr. rer. nat. Stefan Guthe (Graphics Capture and Massively Parallel Computing group, TU Darmstadt).

As already mention, the study of the database revealed systematic differences between individuals, which seem to be related to personality (and maybe mood). Nevertheless, all the findings derived from the experiments in this thesis are bounded to cultural dependencies. The participants for all the experiments were German, while we had both German and Spanish people recorded for our stimuli (MPI and MoCap DBs respectively). In order to make our findings more general, is necessary to study the intercultural differences (is there a subset of the individual differences that is common to members of a given culture) in both the perception of the expressions and the perceived personality. This work is progressing in cooperation with Prof. Dr. Christian Wallraven, Korea University Seoul.

A specific mapping between the Semantic Space and facial *motions* must still be found and it would seem to be appropriate to use a spatiotemporal description of facial expressions. As already mentioned, the abstraction of the Semantic Space makes possible to move from a discrete collection of data to a continuous metric space, allowing us to generate any point on the space from the sampled points. Part III was a proof of concept for this motion synthesis. Nevertheless, the approach used in this part, was rather inelegant, since it requires using all motion in all of the recordings, rather than the relevant, meaning-carrying motions. Thus, we should find the structures that define specific emotions. This would allows us to use or more refined combination of recording elements in order to produce novel expressions. Note that this step should be incorporated in the pipeline before the motion capture trajectories are analyzed, in order to provide an

empirical basis for determining what aspects of the videos are important or and what aspect of the expressions meaning they might carry.

Among the multimodality of facial communication, this thesis focused on the visual channel, more specifically in the temporal information conveyed through it. But, when designing an ECA, animating it is not enough. It is obvious that the static information coming from the same channel or from other channels, as the acoustic one, shall be carefully considered. As shown in Part IV, we have already begun to examine how facial motion and acoustic information are combined in the perception of expressions and personality. Thus, we still require a proper characterization of the audio structure of the founded Semantic Space for emotions and expressions, an equivalent characterization for the effects of visual stylization (control of the appearance) on the aforesaid perception of expressions and personality and, most critically, a study on the interaction of all communication channels. Hence, our lab is already exploring both unimodal characterizations under the pertinent projects leaded by Martin Schorradt and Philipp Hahn. Given the first fruits of these projects [6, 8], we firmly believe that the methodological approach presented in this thesis is well worth to be considered towards finding these new characterizations of the Semantic Space.

Last, but not least, we have begun to explore incorporating the personality and emotional animations presented in this thesis in a new State-of-the-Art ECA. This work will most likely be performed in a large-scale cooperation with a number of professors from the Brandenburg University of Technology, University of Bamberg, and the University of Marburg.

BIBLIOGRAPHY

[1] Susana Castillo, Christian Wallraven, and Douglas W. Cunningham. The semantic space for facial communication. *Computer Animation and Virtual Worlds*, 25(3-4):225–233, 2014.

[2] Susana Castillo, Katharina Legde, and Douglas W. Cunningham. The semantic space for motion-captured facial expressions. *Computer Animation and Virtual Worlds*, 29(3-4):e1823, 2018.

[3] Katharina Legde, Susana Castillo, and Douglas W. Cunningham. Multimodal affect: Perceptually evaluating an affective talking head. *ACM Transactions on Applied Perception (TAP)*, 12(4):17:1–17:17, 2015.

[4] Susana Castillo, Philipp Hahn, Katharina Legde, and Douglas W Cunningham. Personality analysis of embodied conversational agents. In *IVA'18: International Conference on Intelligent Virtual Agents (IVA '18)*, New York, NY, USA, November 5-8 2018. ACM.

[5] Susana Castillo, Douglas W. Cunningham, Christian Winger, and Michael Breuß. Morphological amoeba-based patches for exemplar-based inpainting. In *Journal of WSCG*, volume 26:2, pages 112–121. ISSN 1213-6972, May 2018.

[6] Martin Schorradt, Susana Castillo, and Douglas W. Cunningham. The semantic space for emotional speech and the influence of different methods for prosody isolation on its perception. In *Proceedings of the 15th ACM Symposium on Applied Perception*, SAP '18, pages 15:1–15:8, New York, NY, USA, 2018. ACM.

[7] Katharina Legde, Susana Castillo, and Douglas W Cunningham. Age regression: Rejuvenating 3d-facial scans. In *Short Papers Proceedings WSCG'2018 - 26. Inter-*

national Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2018, pages 190–199. Computer Science Research Notes [CSRN 2802] ISSN 2464-4617, May 2018.

[8] Philipp Hahn, Susana Castillo, and Douglas W. Cunningham. Look me in the lines: The impact of stylization on the recognition of expressions and perceived personality. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, IVA '18, pages 339–340, New York, NY, USA, 2018. ACM.

[9] Martin Schorradt, Katharina Legde, Susana Castillo, and Douglas W. Cunningham. Integration and evaluation of emotion in an articulatory speech synthesis system. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, SAP '15, Tübingen, Germany, September 13-14, 2015*, page 137, 2015.

[10] Douglas W. Cunningham, Mario Kleiner, Christian Wallraven, and Heinrich H. Bülthoff. Manipulating video sequences to determine the components of conversational facial expressions. *ACM Transactions on Applied Perception (TAP)*, 2(3):251–269, jul 2005.

[11] A. Mehrabian. Communication without words. *Psychology Today*, 2:53–55, 1968.

[12] P. Carrera-Levillain and J. Fernandez-Dols. Neutral faces in context: Their emotional meaning and their function. *Journal of Nonverbal Behavior*, 18:281–299, 1994.

[13] Paul Watzlawick, Janet H. Beavin, and Don D. Jackson. *Menschliche Kommunikation: Formen, Störungen, Paradoxien, 10.* Bern u.a., Hans Huber, 1972.

[14] P. Ekman. Universal and cultural differences in facial expressions of emotion. In J. R. Cole, editor, *Nebraska Symposium on Motoivation 1971*, pages 207–283. University of Nebraska Press, Lincoln, NE, 1972.

[15] Emiel Krahmer, Zsofia Ruttkay, Marc Swerts, and Wieger Wesselink. Pitch, eyebrows and the perception of focus. In *In Symposium on Speech Prosody*, 2002.

[16] W. S. Condon and W. D. Ogston. Sound film analysis of normal and pathological behaviour patterns. *Journal of Nervous and Mental Disease*, 143:338–347, 1966.

[17] V. H. Yngve. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567–578. Chicago Linguistic Society, Chicago, 1970.

[18] Johnny R.J. Fontaine, Klaus R. Scherer, Etienne B. Roesch, and Phoebe C. Ellsworth. The world of emotions is not two-dimensional. *Psychological Science*, 18(12):1050–1057, European Conference on Visual Perception (ECVP).

[19] Rensis Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):1–55, 1932.

[20] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum. *The measurement of meaning*. Urbana, USA: University of Illinois Press., 1957.

[21] Douglas Cunningham and Christian Wallraven. *Experimental Design: From User Studies to Psychophysics*. A. K. Peters, Ltd., Natick, MA, USA, 1st edition, 2011.

[22] J.P. de Ruiter, S. Rossignol, L. Vuurpijl, D.W. Cunningham, and W. J. M. Levelt. Slot: A research platform for investigating multimodal communication. *Behavior Research Methods, Instruments & Computers*, 35(3):408–419, 2003.

[23] D. Archer and R. M. Akert. Words and everything else: Verbal and nonverbal cues to social interpretation. *Journal of Personality & Social Psychology*, 35:443–449, 1977.

[24] J. Fernandez-Dols, H. Wallbott, and F. Sanchez. Emotion category accessibility and the decoding of emotion from facial expression and context. *Journal of Nonverbal Behavior*, 15:107–124, 1991.

[25] A. Mehrabian and S. Ferris. Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, 31:248–252, 1967.

[26] Eugene Hsu, Kari Pulli, and Jovan Popović. Style translation for human motion. *ACM Transactions on Graphics (ToG)*, 24(3):1082–1089, July 2005.

[27] John Funge, Xiaoyuan Tu, and Demetri Terzopoulos. Cognitive modeling: Knowl-

edge, reasoning and planning for intelligent characters. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pages 29–38, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[28] Mubbasir Kapadia, Alexander Shoulson, Funda Durupinar, and Norman I Badler. Authoring multi-actor behaviors in crowds with diverse personalities. In *Modeling, Simulation and Visual Analysis of Crowds*, pages 147–180. Springer New York, 2013.

[29] Jan Allbeck and Norman Badler. Toward representing agent behaviors modified by personality and emotion. *Embodied Conversational Agents at the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2:15–19, 2002.

[30] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 173–182, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[31] J M Digman. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1):417–440, 1990.

[32] Vinay Bettadapura. Face expression recognition and analysis: The state of the art. *Computing Research Repository (CoRR)*, abs/1203.6722, 2012.

[33] Junghyun Ahn, Stéphane Gobron, Quentin Silvestre, and Daniel Thalmann. Asymmetrical Facial Expressions based on an Advanced Interpretation of Two-dimensional Russells Emotional Model. In *proceedings of ENGAGE 2010, Zermatt, Switzerland, September 13-15, 2010*, 2010.

[34] Haishan Chen, Huailin Dong, and Bochao Hu. Research on emotion modeling based on three-dimension emotional space. In *Communications and Information Processing*, pages 310–319. Springer, 2012.

[35] Kathrin Kaulard, Douglas W. Cunningham, Heinrich H. Bülthoff, and Christian Wallraven. The mpi facial expression database — a validated database of emotional

and conversational facial expressions. *PLoS ONE*, 7(3):e32321, 03 2012.

[36] Evgeni N Sokolov and Wolfram Boucsein. A psychophysiological model of emotion space. *Integrative Physiological and Behavioral Science*, 35(2):81–119, 2000.

[37] James A Russell. A circumplex model of affect. *Journal of Personality & Social Psychology*, 39(6):1161–1178, 1980.

[38] M. S. M. Yik, J. A. Russell, and L. F. Barrett. Structure of self-reported current affect: Integration and beyond. *Journal of Personality & Social Psychology*, 77(3):600–619, 1999.

[39] Lisa Feldman Barrett and James A. Russell. The structure of current affect controversies and emerging consensus. *Current Directions in Psychological Science*, 8(1):10–14, 1999.

[40] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436, 1997.

[41] D. G. Pelli. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437–442, 1997.

[42] M. Kleiner, D. Brainard, and D. Pelli. What's new in psychtoolbox-3? *Perception 36, European Conference on Visual Perception (ECVP) Abstract Supplement*, European Conference on Visual Perception (ECVP).

[43] J Kevin Ford, Robert C MacCallum, and Marianne Tait. The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(2):291–314, 1986.

[44] Amy E Hurley, Terri A Scandura, Chester A Schriesheim, Michael T Brannick, Anson Seers, Robert J Vandenberg, and Larry J Williams. Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, 18(6):667–683, 1997.

[45] Pedro R Peres-Neto and Donald A Jackson. How well do multivariate data sets match? the advantages of a procrustean superimposition approach over the mantel

test. *Oecologia*, 129(2):169–178, 2001.

[46] Eleanor Rosch. *, in Eleanor Rosch and Barbara B. Lloyd, eds.: Cognition and Categorization.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.

[47] Douglas W. Cunningham and Christian Wallraven. Dynamic information for the recognition of conversational expressions. *Journal of Vision*, 9(13):1 – 17, 2009.

[48] Phoebe C Ellsworth and Linda M Ludwig. Visual behavior in social interaction. *Journal of Communication*, 22(4):375–403, 1972.

[49] Cheongjae Lee and Gary Geunbae Lee. Emotion recognition for affective user interfaces using natural language dialogs. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 798–801, Aug 2007.

[50] Byron Reeves and Clifford Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, New York, NY, USA, 1996.

[51] Zhigang Deng, Jeremy Bailenson, John P Lewis, and Ulrich Neumann. Perceiving visual emotions with speech. In *International Workshop on Intelligent Virtual Agents*, pages 107–120. Springer, 2006.

[52] Martin Breidt, Christian Wallraven, D Cunningham, and H Bulthoff. Combining 3d scans and motion capture for realistic facial animation. *Proceedings der Eurograph,(Eds.) Julian and F. and P. Cano and The Eurographics Association*, pages 63–66, 2003.

[53] Douglas W. Cunningham, Mario Kleiner, Heirich H. Bülthoff, and Christian Wallraven. The components of conversational facial expressions. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, APGV '04, pages 143–150, New York, NY, USA, 2004. ACM.

[54] Manfred Nusseck, Douglas W. Cunningham, Christian Wallraven, and Heinrich H. Bülthoff. The contribution of different facial regions to the recognition of conversa-

tional expressions. *Journal of Vision*, 8(8):1, 2008.

[55] Stephanie N. Mullins-Sweatt, Janetta E. Jamerson, Douglas B. Samuel, David R. Olson, and Thomas A. Widiger. Psychometric properties of an abbreviated instrument of the five-factor model. *Assessment*, 13(2):119–137, 2006.

[56] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pages 187–194, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[57] Frank E Pollick, Harold Hill, Andrew Calder, and Helena Paterson. Recognising facial expression from spatially and temporally modified movements. *Perception*, 32(7):813–826, 2003. PMID: 12974567.

[58] Christian Wallraven, Martin Breidt, Douglas W. Cunningham, and Heinrich H. Bülthoff. Evaluating the perceptual realism of animated facial expressions. *ACM Transactions on Applied Perception (TAP)*, 4(4):4:1–4:20, February 2008.

[59] Douglas W. Cunningham and Christian Wallraven. The interaction between motion and form in expression recognition. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, APGV '09, pages 41–44, New York, NY, USA, 2009. ACM.

[60] Eva G. Krumhuber, Arvid Kappas, and Antony S. R. Manstead. Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, 5(1):41–46, 2013.

[61] Martin A. Giese and Tomaso Poggio. Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision*, 38(1):59–73, Jun 2000.

[62] Nikolaus F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5):2–2, 09 2002.

[63] Justine Cassell, Joseph Sullivan, and Elizabeth Churchill. *Embodied Conversational Agents*. MIT Press Cambridge, Massachusetts, London, England, 2000.

[64] Lawrence A Pervin. *The science of personality.* New York : Wiley & Sons, 1996.

[65] Hans J Eysenck. *Dimensions of personality*, volume 5. Transaction Publishers, 1950.

[66] Raymond B. Cattell. The Scientific Analysis of Personality. Baltimore, Md. Penguin Books, Inc., 1965, 399 p. *Psychology in the Schools*, 3(1):93–93, 1966.

[67] Heather EP Cattell and Alan D Mead. The sixteen personality factor questionnaire (16pf). *The SAGE handbook of personality theory and assessment*, 2:135–178, 2008.

[68] Christopher J. Soto and Joshua J. Jackson. Five-factor model of personality. *Psychology*, February 2013.

[69] J. M. Digman. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, 41(1):417–440, 1990.

[70] Robert R. McCrae and Oliver P. John. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60(2):175–215, 1992.

[71] Clifford Nass, Youngme Moon, BJ Fogg, Byron Reeves, and Chris Dryer. Can computer personalities be human personalities? In *Conference companion on Human factors in computing systems*, pages 228–229. ACM, 1995.

[72] V. Vinayagamoorthy, M. Gillies, A. Steed, E. Tanguy, X. Pan, C. Loscos, and M. Slater. Building Expression into Virtual Characters. In Brian Wyvill and Alexander Wilkie, editors, *Eurographics 2006 - State of the Art Reports*. The Eurographics Association, 2006.

[73] Angelo Cafaro, Hannes Högni Vilhjálmsson, Timothy Bickmore, Dirk Heylen, Kamilla Rún Jóhannsdóttir, and Gunnar Steinn Valgardhsson. *First Impressions: Users' Judgments of Virtual Agents' Personality and Interpersonal Attitude in First Encounters*, pages 67–80. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[74] Kerstin Ruhland, Katja Zibrek, and Rachel McDonnell. Perception of personality through eye gaze of realistic and cartoon models. In *Proceedings of the ACM*

SIGGRAPH Symposium on Applied Perception, pages 19–23. ACM, 2015.

[75] Clifford Nass, Katherine . Isbister, and Eun-Ju Lee. Truth is beauty: Researching embodied conversational agents. In *Embodied Conversational Agents*, pages 374–402. MIT Press, Cambridge, MA, USA, 2000.

[76] Elisabetta Bevacqua, Etienne De Sevin, Sylwia Julia Hyniewska, and Catherine Pelachaud. A listener model: introducing personality traits. *Journal on Multimodal User Interfaces*, 6(1-2):27–38, 2012.

[77] Jennifer Hyde, Elizabeth J Carter, Sara Kiesler, and Jessica K Hodgins. Perceptual effects of damped and exaggerated facial motion in animated characters. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6. IEEE, 2013.

[78] Katja Zibrek and Rachel McDonnell. Does render style affect perception of personality in virtual humans? *Proceedings of the ACM Symposium on Applied Perception*, 2014:111–115, 2014.

[79] Marc Schröder. The semaine api: Towards a standards-based framework for building emotion-oriented systems. *Advances in Human-Computer Interaction*, 2010, January 2010.

[80] Marc Schröder, Roddy Cowie, Dirk K.J. Heylen, Maja Pantic, Catherine Pelachaud, and Björn Schuller. Towards responsive sensitive artificial listeners. In *Proceedings of the Fourth International Workshop on Human-Computer Conversation*, number 2008/16200, United Kingdom, 10 2008. University of Sheffield.

[81] M. Schröder, E. Bevacqua, F. Eyben, H. Gunes, D. Heylen, M. ter Maat, S. Pammi1, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. Valstar, and M. Wöllmer. A demonstration of audiovisual sensitive artificial listeners. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–2, Sept 2009.

[82] Elisabetta Bevacqua, Satish Pammi, Catherine Pelachaud, Marc Schröder, and Etienne de Sevin. D5b SAL multimodal generation component optimised for real-

time behaviour. *SEMAINE the sensitive agent project*, 24 September 2010.

[83] Ellen Douglas-Cowie, Roddy Cowie, Cate Cox, Noam Amier, and Dirk K. J. Heylen. The sensitive artificial listener: an induction technique for generating emotionally coloured conversation. In L. Devillers, J-C. Martin, R. Cowie, E. Douglas-Cowie, and A. Batliner, editors, *LREC Workshop on Corpora for Research on Emotion and Affect*, number WP 08-02, pages 1–4, Paris, France, 2008. ELRA.

[84] Elisabetta Bevacqua, Sathish Pammi, Sylwia Julia Hyniewska, Marc Schröder, and Catherine Pelachaud. Multimodal backchannels for embodied conversational agents. In *Intelligent Virtual Agents*, pages 194–200, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[85] Florian Eyben, Hatice Gunes, Maja Pantic, Marc Schröder, Björn Schuller, Michel F. Valstar, and Martin Wöllmer. D3c User-profiled human behaviour interpreter. *SEMAINE the sensitive agent project*, 24 September 2010.

[86] Etienne de Sevin, Sylwia Julia Hyniewska, and Catherine Pelachaud. Influence of personality traits on backchannel selection. In *Intelligent Virtual Agents*, pages 187–193, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[87] A. N. Oppenheim. *Questionnaire Design, Interviewing, and Attitude Measurement.* Continuum, 1992.

[88] Paul T Costa Jr and Robert R McCrae. Domains and facets: Hierarchical personality assessment using the revised neo personality inventory. *Journal of Personality Assessment*, 64(1):21–50, 1995.

[89] Zsófia Ruttkay, Claire Dormann, and Han Noot. Embodied conversational agents on a common ground. In Zsófia Ruttkay and Catherine Pelachaud, editors, *From Brows to Trust*, pages 27–66. Kluwer Academic Publishers, Norwell, MA, USA, 2004.

[90] M. Slater, D.-P. Pertaub, and A. Steed. Public speaking in virtual reality: facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9, March 1999.

[91] J. Klein, Y. Moon, and R.W. Picard. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14(2):119–140, 02 2002.

[92] Kristine L. Nowak and Frank Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494, 10 2003.

[93] Michael Argyle and Peter Trower. *Person to Person: Ways of Communicating.* Harper and Row Publishers, 1979.

[94] Albert Mehrabian. *Silent Messages*, volume first ed. Wadsworth Publishing Company, January 1971.

[95] Albert Mehrabian and John T Friar. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology*, 33(3):330–336, 1969.

APPENDICES

EXPRESSIONS

In this thesis I gathered a new database of conversational facial expressions both in video and Motion-Capture data formats. A total of ten Spanish people, five females and five males (without any previous experience in acting) were recorded while performing 62 different facial expressions following the technique called "method acting protocol" proposed by Kaulard et al. et al. in [35]. These 62 expressions were a corrected and extended version of the 55 expressions video-database previously proposed by Kaulard et al. [35]. This Appendix contains exemplar scenarios used for the recording of these expressions.

A.1 Exemplar scenarios for the Recorded Expressions

In the following we provide a list of the recorded expressions in our database, together with the categorization of each expression and a scenario in English that could serve as an example for the ones used to trigger the desired facial expression on the actor during the recordings.

Expression		Identifier	Scenario	Expression Type
Agree	Considered	`[AgrCons]`	Someone suggests to try something. You hesitate first but then you consent	***Conversational Expressions***
	Continue	`[AgrCont]`	During a conversation you signal your partner that you have understood everything and that s/he can keep on talking	***Conversational Expressions***
	Pure	`[Agr]`	You share someone's verbalized opinion	***Conversational Expressions***
	Reluctant	`[AgrRel]`	Someone suggests to try something. You consent even when you have some concerns about it	***Conversational Expressions***
Aha, Right		`[Aha]`	Now I get it!	***Conversational Expressions***
Disagree	Pure	`[Disa]`	You do not share someone's verbalized opinion	***Conversational Expressions***

Expression		Identifier	Scenario	Expression Type
	Reluctant	[DisaRel]	Someone suggests to try something. You are not fully up for it and finally decline	***Conversational Expressions***
	Considered	[DisaCons]	Someone suggests to try something. You hesitate first and finally decide not to go for it	***Conversational Expressions***
Anger		[Ang]	Your less favorite flat mate has taken your dinner out of the fridge which you were looking forward to eat all day long	***Basic Emotional***
Disgust		[Disg]	You find molded food in your fridge after you come home from a journey	***Basic Emotional***
Fear	"Oh my God!"	[FeOMG]	After leaving your flat you realize you forgot to switch off the cooker	***Subordinate Emotional***

Expression		Identifier	Scenario	Expression Type
	Terror	[FeTe]	There is a tarantula climbing your back	***Basic Emotional***
Happy	Laughing	[HapLau]	You are laughing about a joke	***Basic Emotional***
	Achivement	[HapAch]	You have reached a goal and you are happy to have finished it	***Subordinate Emotional***
	Satiated	[HapSat]	You are lying on your couch after a delicious dinner	***Extended Emotional*** (Includes Satisfaction)
	SchadenFreude	[HapSF]	Someone whom you don't like slips on a banana peel in front of you	***Subordinate Emotional***
Sadness		[Sad]	Someone close to you has passed away	***Basic Emotional***
Surprise	Neutral	[Surp=]	You are going to grab a tool from a table to discover that is just a painting	***Basic Emotional***
	Pleasant Surprise	[Surp+]	You find a 50€ bill in your pocket	***Basic Emotional***
	Unpleasant Surprise	[Surp-]	You realized your lost your wallet	***Basic Emotional***

Expression		**Identifier**	**Scenario**	**Expression Type**
Contempt		`[Cont]`	You think of someone you despise	***Extended Emotional*** (Related to Disgust)
Arrogant		`[Arrog]`	Only you are the best!	***Extended Emotional*** (Includes Contempt)
Embarrassment		`[Emba]`	Your pants rip off when you bend down to pick something	***Extended Emotional***
Evasive		`[Evas]`	Your colleague asks about your opinion on a haircut you find terrible	***Extended Emotional***
Pain	Felt	`[PainF]`	While doing sports you have an accident suddenly in which you wrench one ankle and graze your knee	***Subordinate Emotional***
	Seen	`[PainS]`	You watch a TV transmission an sport event. Suddenly one player has a serious accident. You can see bones sticking out of the player's body	***Subordinate Emotional***

Expression		Identifier	Scenario	Expression Type
Smile	Endearment	`[SmlEnd]`	A little girl smiles at you	***Subordinate Emotional***
	Encouraging	`[SmlEnc]`	Someone is worried, you tell him: "Cheer up, everything will work!"	***Subordinate Emotional***
	Flirting	`[SmlFli]`	An attractive individual makes eye contact with you from the distance and you try to seduce him/her	***Subordinate Emotional***
	Reluctant	`[SmlRel]`	A friend of you wants to go out this evening. You do not want to go with him/her. S/He tells you that a couple of other friends will also be around who you want to see again	***Subordinate Emotional***
	Sardonic	`[SmlSar]`	You said a sardonic joke concerning one of your friends	***Subordinate Emotional***

Expression	Identifier	Scenario	Expression Type
Sad/Nostalgia	[SmlSN]	You recall a pleasant situation that happened to you in the past and get the feeling that everything was better back then	***Subordinate Emotional***
Triumphant	[SmlTri]	You achieved something which someone didn't believe possible	***Extended Emotional*** (Includes Pride in Achievement)
Uncertain	[SmlUnc]	Someone is excited telling you something that happened and you are not sure if it is positive	***Subordinate Emotional***
Wallace and Gromit	[SmlWG]	You smile like Wallace and Gromit or like in a toothpaste commercial	***Conversational Expressions***
Winning	[SmlWin]	The parents of your girlfriend visit you. You open the door and welcome them	***Subordinate Emotional***

Expression		Identifier	Scenario	Expression Type
	"Yeah, right!"	[SmlYAI]	Someone tells you something incredible and you think: "Yeah, as if..."	***Subordinate Emotional***
Guilty		[Guilt]	Someone is punished because a mistake you did	***Subordinate Emotional***
Relief		[Reli]	You thought you have lost your phone to find out that you just misplaced it	***Subordinate Emotional***
Shame		[Sham]		***Subordinate Emotional***
Bored		[Bor]	You have been waiting on a line for a long time with nothing else to do	***Conversational Expressions***
Annoyed	Bothering	[AnnoyBo]	You have to do tons of work which you do not want to do at all. The night before the deadline you realize that you have to work all night long	***Conversational Expressions***

Expression	Identifier	Scenario	Expression Type
Rolling Eyes	[AnnoyRE]	You explain something for the 10th time and your listener still does not get it	***Conversational Expressions***
Confused	[Conf]	You lose the way in a foreign city	***Conversational Expressions***
"I don't Care!"	[NoCare]	Someone suggests something but you are not interested in it at all	***Conversational Expressions***
"I didn't Hear!"	[NoHear]	Someone talks to you but you cannot understand it because the environment is too loud	***Conversational Expressions***
Disbelief	[Disbe]	Someone tells you a true story, however, you do not want to believe it	***Conversational Expressions***
"I don't Know!" (Clueless)	[NotKnow]	Someone asks you for the name of the Ugandan president	***Conversational Expressions***
"I don't Understand!"	[NoUnd]	Someone talks to you in an unknown language	***Conversational Expressions***

Expression		Identifier	Scenario	Expression Type
Imagine	Negative	`[Img-]`	You imagine something unpleasant in your future	***Conversational Expressions***
	Positive	`[Img+]`	You imagine something pleasant in your future	***Conversational Expressions***
Impressed		`[Impre]`	You observe someone dancing and think: "Wow, that's really good!"	***Conversational Expressions***
Insecurity		`[Insec]`	You use an expensive device at an exposition and suddenly it stops working. You are not sure if this was your fault	***Conversational Expressions***
Compassion		`[Compa]`	Your best friend tells you that s/he has broken up	***Conversational Expressions***
Maybe, Not Convinced		`[Maybe]`	Someone gives a solution to a problem but you are not fully sure it would work	***Conversational Expressions***

Expression		Identifier	Scenario	Expression Type
Thinking	Considering	`[ThCons]`	Someone makes a suggestion and you hesitate	***Conversational Expressions***
	Remember Negative	`[Remb-]`	You recall an awkward situation that happened to you in the past	***Conversational Expressions***
	Remember Neutral	`[Remb=]`	You think about what you had for breakfast one week ago	***Conversational Expressions***
	Remember Positive	`[Remb+]`	You recall a pleasant situation that happened to you in the past	***Conversational Expressions***
	Problem Solving	`[ThPSol]`	You think of how old you are in ... months	***Conversational Expressions***
Tired		`[Tired]`	The only thing you want to do is to lie in the bed after a long working day	***Conversational Expressions***
Treudoof		`[Treud]`	Innocent, "Bambi eyes"	***Conversational Expressions***

Table A.1: The 62 emotions recorded in two different modalities for this thesis. The abbreviations in brackets are used along the figures contained in this thesis for better visualization.

OCEAN QUESTIONNAIRES

For the sake of completeness, the current Appendix contains the English and German versions of the FFMRFs used in our experiments.

B.1 QUESTIONNARIE

Instructions

Please describe the individual on a seven point scale on each of the following 30 personality traits. Please provide a rating for all 30 traits. For example, on the first trait (anxiousness), checking "extremely low" would indicate that you think the individual is extremely low in anxiousness (i.e., relaxed, unconcerned, cool). Checking "neutral" would indicate that you think the individual is neither high nor low in anxiousness (does not differ from the average person) or that you are unable to decide.

Neuroticism versus Emotional Stability:

1. **Anxiousness** (fearful, apprehensive) (relaxed, unconcerned, cool)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

2. **Angry Hostility** (angry, bitter) (even-tempered)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

3. **Depressiveness** (pessimistic, glum) (optimistic)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

4. **Self-consciousness** (timid, embarrassed) (self-assured, glib, shameless)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

5. **Impulsivity** (tempted, urgency) (controlled, restrained)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

6. **Vulnerability** (helpless, fragile) (clear-thinking, fearless, unflappable)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

Extraversion versus Introversion:

8. **Warmth** (cordial, affectionate, attached) (cold, aloof, indifferent)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

9. **Gregariousness** (sociable, outgoing) (withdrawn, isolated)
   ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

10. **Assertiveness** (dominant, forceful) (unassuming, quiet, resigned)
    ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

11. **Activity** (vigorous, energetic, active) (passive, lethargic)
    ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

12. **Excitement-Seeking** (reckless, daring) (cautious, monotonous, dull)
    ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

13. **Positive Emotions** (high-spirited) (placid, anhedonic)
    ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

Openness versus Closedness to one's own Experience:

14. **Fantasy** (dreamer, unrealistic, imaginative) (practical, concrete)
    ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

15. **Aesthetics** (aberrant interests, aesthetic) (uninvolved, no aesthetic interests)
    ☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

16. **Feelings** (self-aware) (unassuming, quiet, resigned)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

17. **Actions** (unconventional, eccentric) (passive, lethargic)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

18. **Ideas** (strange, odd, peculiar, creative) (pragmatic, rigid)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

19. **Values** (permissive, broad-minded) (traditional, inflexible, dogmatic)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

Agreeableness versus Antagonism:

20. **Trust** (gullible, naïve, trusting) (skeptical, cynical, suspicious, paranoid)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

21. **Straightforwardness** (confiding, honest) (cunning, manipulative, deceptive)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

22. **Altruism** (sacrificial, giving) (stingy, selfish, greedy, exploitative)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

23. **Compliance** (docile, cooperative) (oppositional, combative, aggressive)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

24. **Modesty** (meek, self-effacing, humble) (confident, boastful, arrogant)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

25. **Tender-Mindedness** (soft, empathetic) (tough, callous, ruthless)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

Conscientiousness versus Undependability:

26. **Competence** (perfectionistic, efficient) (lax, negligent)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

27. **Order** (ordered, methodical, organized) (haphazard, disorganized, sloppy)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

28. **Dutifulness** (rigid, reliable, dependable) (casual, undependable, unethical)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

29. **Achievement** (workaholic, ambitious) (aimless, desultory)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

30. **Self-Discipline** (dogged, devoted) (hedonistic, negligent)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

31. **Deliberation** (cautious, ruminative, reflective) (hasty, careless, rash)
☐ extremely high ☐ high ☐ somewhat high ☐ neutral ☐ somewhat low ☐ low ☐ extremely low

B.2 Fragebogen

Anleitung

Bitte beschreiben Sie den gezeigten Charakter auf einer sieben-Punkt-Skala bezüglich der folgenden 30 Persönlichkeitseigenschaften. (Bitte bewerten Sie jede der 30 Eigenschaften.) Für die erste Eigenschaft (Ängstlichkeit) beispielsweise würde "Extrem Niedrig" für einen Charakter sprechen, der sehr wenig Ängstlichkeit aufweist (d.h. entspannt, unbekümmert wirkt). Bei der Auswahl "Neutral" ist kein erhöhterverminderter Grad an Ängstlichkeit zu erkennen oder es ist schwer festzustellen. Markieren Sie die Kästchen, welche auf den Charakter zutreffen für jede der 30 Eigenschaften.

Neurotizismus (Neuroticism versus Emotional Stability)

1. **Ängstlichkeit** (ängstlich, besorgt) (entspannt, unbekümmert)
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

2. **Feindlichkeit** (wütend, verbittert) (gelassen)
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

3. **Depressivität** (pessimistisch, deprimiert) (optimistisch)
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

4. **Selbstbewusstsein** (zaghaft, verlegen) (selbstsicher, wortgewandt, schamlos)
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

5. **Impulsivität** (verleitet sein, dringlich, spontan) (kontrolliert, verhalten)
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

6. **Verletzlichkeit** (hilflos, zerbrechlich) (klar denkend, angstfrei, unerschütterlich)
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

Extraversion und Introversion (Extraversion versus Introversion)

8. **Herzlichkeit** (freundlich, zugeneigt, anhänglich) (kalt, distanziert, gleichgültig))
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

9. **Geselligkeit** (kontaktfreudig, extrovertiert) (zurückgezogen, introvertiert)
   ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

10. **Bestimmtheit** (dominant, energisch) (anspruchslos, ruhig, gleichgültig)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

11. **Lebhaftigkeit** (lebendig, tatkräftig, aktiv) (passiv, träge)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

12. **Risikobereitschaft** (rücksichtslos, wagemutig) (achtsam, abwechslungslos, langweilig)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

13. **Positive Emotionen** (begeistert) (gelassen, ohne Vergnügen)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

Offenheit für Erfahrungen (Openness versus Closedness)

14. **Fantasie** (Träumer, unrealistisch, einfallsreich) (praktisch, konkret)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

15. **Ästhetik** (unkonventionelle Interessen, ästhetisch) (unbetroffen, keine ästhetischen Interessen)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

16. **Gefühle** (seiner selbst bewusst) (beeinträchtigt, nichts ahnend, gefühlsblind)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

17. **Handlungen** (unkonventionell, exzentrisch) (passiv, träge, gewohnheitsmäßig)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

18. **Ideen** (seltsam, merkwürdig, eigen, kreativ) (pragmatisch, starr)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

19. **Werte** (tolerant, aufgeschlossen) (konservativ, unnachgiebig, rechthaberisch)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

Verträglichkeit (Agreeableness versus Antagonism)

20. **Vertrauen** (leichtgläubig, naiv, gutgläubig) (skeptisch, zynisch, argwöhnisch, paranoid)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

21. **Direktheit** (anvertrauend, ehrlich) (raffiniert, manipulierend, täuschend)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

22. **Selbstlosigkeit** (aufopfernd, großzügig) (geizig, selbstsüchtig, gierig, ausbeuterisch)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

23. **Nachgiebigkeit** (fügsam, kooperierend) (gegensätzlich, streitlustig, aggressiv)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

24. **Bescheidenheit** (demütig, zurückhaltend, bescheiden) (selbstsicher, überheblich, arrogant)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

25. **Empfindsamkeit** (sanft, mitfühlend) (hart, gefühllos, rücksichtslos)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

Gewissenhaftigkeit (Conscientiousness versus Undependability)

26. **Kompetenz** (perfektionistisch, effizient) (nachlässig, fahrlässig)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

27. **Organisation** (geordnet, methodisch, organisiert) (planlos, ungeordnet, schlampig)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

28. **Pflichtbewusstsein** (unnachgiebig, vertrauenswürdig, zuverlässig) (sorglos, unzuverlässig, unethisch)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

29. **Leistung** (Workaholic, ehrgeizig) (ziellos, halbherzig)
☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

30. **Selbstdisziplin** (hartnäckig, hingebungsvoll) (vergnügungssüchtig, fahrlässig)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig

31. **Bedächtigkeit** (achtsam, nachdenklich, reflektierend) (voreilig, unvorsichtig, unüberlegt)
    ☐ Extrem Hoch ☐ Hoch ☐ Etwas Hoch ☐ Neutral ☐ Etwas Niedrig ☐ Niedrig ☐ Extrem Niedrig