CORRELATION BETWEEN COMPUTER RECOGNIZED FACIAL EMOTIONS AND INFORMED EMOTIONS DURING A CASINO COMPUTER GAME

by

Nils Reichert

Bachelor of Science, BRSU, 2009

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

Master of Science in the Graduate Academic Unit of Computer Science

Supervisor: Kenneth B. Kent, Ph.D., Computer Science
Rainer Herpers, Ph.D., Computer Science

Examining Board: Gerhard Dueck, Ph.D., Computer Science
Przemyslaw Pochec, Ph.D., Computer Science
Eduardo Castillo-Guerra, Ph.D., Electrical and Computer Engineering

This thesis is accepted by the Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

September, 2012

©Nils Reichert, 2012
ABSTRACT

Emotions play an important role for everyday communication. Different methods allow computers to recognize emotions. Most are trained with acted emotions and it is unknown if such a model would work for recognizing naturally appearing emotions. An experiment was setup to estimate the recognition accuracy of the emotion recognition software SHORE, which could detect the emotions angry, happy, sad, and surprised. Subjects played a casino game while being recorded. The software recognition was correlated with the recognition of ten human observers. The results showed a strong recognition for happy, medium recognition for surprised, and a weak recognition for sad and angry faces. In addition, questionnaires containing self-informed emotions were compared with the computer recognition, but only weak correlations were found. SHORE was able to recognize emotions almost as well as humans were, but if humans had problems to recognize an emotion, then the accuracy of the software was much lower.
ACKNOWLEDGEMENTS

Thanks to my advisors Prof. Kenneth B. Kent (University of New Brunswick, Canada) and Prof. Rainer Herpers (Bonn-Rhein-Sieg University of Applied Sciences, Germany) for their support and suggestions.
# Table of Contents

ABSTRACT ...................................................................................................................... ii

ACKNOWLEDGEMENTS ............................................................................................. iii

Table of Contents ............................................................................................................. iv

List of Tables .................................................................................................................... vi

List of Figures ................................................................................................................ viii

1 Introduction and Purpose ........................................................................................... 1

2 Theoretical Background ............................................................................................ 9

  2.1 Emotion .................................................................................................................. 9
  2.1.1 Emotion-Research in the Past .................................................................... 10
  2.1.2 Emotion Recognition ................................................................................. 17
  2.1.3 Other Aspects of Emotions ........................................................................ 17

  2.2 Data Mining ......................................................................................................... 20

  2.3 Machine Learning ............................................................................................... 20

  2.4 Computer Based Emotion Recognition ........................................................... 21

  2.5 Evaluation with Questionnaires ........................................................................ 23

  2.6 Recent Emotion Recognition Research ............................................................ 25

3 Previous Work ......................................................................................................... 27

  3.1 Previous BRSU Report ........................................................................................ 27
  3.1.1 Emotion recognizing methods .................................................................... 27
  3.1.1.1 Acoustic .............................................................................................. 27
  3.1.1.2 Visual .................................................................................................. 32
  3.1.1.3 Other sources ...................................................................................... 40
  3.1.2 Combining specific features ...................................................................... 42
  3.1.3 Conclusion ................................................................................................. 43

  3.2 Preliminary Research and Development Work .................................................... 45
  3.2.1 wxWidgets ................................................................................................. 45
  3.2.2 Implementation .......................................................................................... 47
  3.2.3 Results Verification and Validation ........................................................... 61
  3.2.4 Conclusion and Future Work ..................................................................... 72

4 Materials and Methods ............................................................................................ 75

  4.1 Description of Experimental Setup ...................................................................... 76

  4.2 Instrumentation ................................................................................................... 82
  4.2.1 Computer-Based Image Processing ........................................................... 82
  4.2.1.1 Extraction of Images - Video Processing ........................................... 82
  4.2.1.2 Conversion of Image Format- Image Processing ............................... 83
  4.2.1.3 Emotion Extraction - Face Detection and Emotion Recognition ...... 83
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1.4</td>
<td>SHORE Library Utilization for Analysis Purpose</td>
<td>90</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Data Recording for Emotion Analysis</td>
<td>91</td>
</tr>
<tr>
<td>4.2.2.1</td>
<td>Casino Game Machine Score</td>
<td>91</td>
</tr>
<tr>
<td>4.2.2.2</td>
<td>Emotion Analysis</td>
<td>92</td>
</tr>
<tr>
<td>4.2.2.3</td>
<td>Other Actions and Reactions of the Subject</td>
<td>93</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Alternative Emotion Retrieval</td>
<td>94</td>
</tr>
<tr>
<td>4.2.3.1</td>
<td>Digital Questionnaire</td>
<td>94</td>
</tr>
<tr>
<td>4.2.3.2</td>
<td>Paper-Based Questionnaire</td>
<td>95</td>
</tr>
<tr>
<td>4.2.3.3</td>
<td>Compare Emotions with Emotional Events</td>
<td>96</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Generation of Reference Ground-Truth</td>
<td>97</td>
</tr>
<tr>
<td>4.3</td>
<td>Data Collection</td>
<td>98</td>
</tr>
<tr>
<td>4.3.1</td>
<td>SHORE-Based Console Application</td>
<td>101</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Moodmeter</td>
<td>106</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Automation Scripts</td>
<td>108</td>
</tr>
<tr>
<td>4.4</td>
<td>Sampling</td>
<td>110</td>
</tr>
<tr>
<td>5</td>
<td>Experimental Analysis and Findings</td>
<td>114</td>
</tr>
<tr>
<td>5.1</td>
<td>Computer-Based Emotion Analysis</td>
<td>117</td>
</tr>
<tr>
<td>5.2</td>
<td>Computer-Based versus Self-Reported Emotions</td>
<td>123</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions and Future Work</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>Appendix A (Thesis)</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Appendix B (UNB R&amp;D)</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Curriculum Vitae</td>
<td></td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Confusion matrix for speaker-dependent emotion recognition, from [89]. The emotion on the left is the real emotion, percentages show as which emotion it is detected. The reason for a total sum in each row slightly below 100% has its cause in rounding down to one digit............................................................................................................. 31

Table 2: Confusion matrix for speaker-independent emotion recognition, from [89].... 31

Table 3: Emotion analysis test results for facial-independent emotion recognition, from [74] using tenfold cross-validation with 340 samples........................................................................ 36

Table 4: Confusion matrix for face-dependant emotion recognition, from [106]........... 37

Table 5: Confusion matrix for face-independent emotion recognition, from [106]...... 37

Table 6: Change of visual emotion recognition when adding acoustic features, from [106]................................................................................................................................ 43

Table 7: Change of the acoustic recognition when adding visual features, from [119].. 43

Table 8: Parameter in the Label control........................................................................ 47

Table 9: Grouped classes that are serialized with a custom serialization ...................... 48

Table 10: Title of asked people that were asked to evaluate the editor ....................... 64

Table 11: Parameters for the SHORE Engine, names and descriptions summarized from [137] The reason for not activating to analyse the other parameters is explained on last paragraph! ....................................................................................................................... 85

Table 12: Example content of the file stats.log............................................................. 102

Table 13: Example extract of the created emotions list............................................... 103

Table 14: Example for EZK answers............................................................................ 107
List of Figures

Figure 1: Actors and connection with Game Machines ................................................................. 5
Figure 2: Valve's arousal measurement by observing skin conductance, from [18] .............. 7
Figure 3: The Four Temperaments, from [31] ......................................................................... 11
Figure 4: Photographs illustrating emotions, from [34] ............................................................ 13
Figure 5: Ekman’s seven basic emotions, expressed by Tim Roth, from [45] ................. 16
Figure 6: Classification of closed-ended questionnaire items, from [76 p. 224] ............ 24
Figure 7: Conclude emotion by frequency with frequency range of meaningful signal, from [86] .................................................................................................................................. 29
Figure 8: 20 calculated Eigen-FFT, from [90] ...................................................................... 30
Figure 9: Importance of each fiducial point according to sensitivity analysis. The importance is illustrated by the size of a point, from [98] ................................................................. 33
Figure 10: Facial motion measurements, from [70] ............................................................... 34
Figure 11: „The learned TAN structure for the facial features”, from [70] ...................... 34
Figure 12: Gabor wavelets and their coverage of the spatial frequency plane, from [105] ........................................................................................................................................ 35
Figure 13: Comparison between accuracy for person learned and foreign person emotion recognition, from [70] ........................................................................................................ 36
Figure 14: Trough inflexion three dimensions are enough, from [102] ............................ 39
Figure 15: Alignment of the embedding coordinates of multiple subjects, from [108] .... 40
Figure 16: Different expressions of happiness: Disingenuous, accretive, caress and doleful, from [120] ........................................................................................................ 44
Figure 17: Currently used XML structure for screens in game machines. Some areas have been replaced with "[...]" to give a better overview.

Figure 18: XML structure to show a Label control on a screen.

Figure 19: The class EventManagement as example for Serialization.

Figure 20: The class EventData to manage existing game machine events.

Figure 21: Classes to define game machine states and to group them.

Figure 22: The different functionalities for the object Question.

Figure 23: The class Questionnaire.

Figure 24: Assigning and filtering with Categories. Left form displays the assigned Categories of the selected Question in the right form.

Figure 25: Classes to offer a defineable hierarchy.

Figure 26: Classes for the user management.

Figure 27: Excerpt from the stored resource XMLBase.

Figure 28: Except from the stored resource XMLLabel.

Figure 29: Classes for exporting a questionnaire to an XML-file.

Figure 30: Excerpt from Visual Studios code coverage of the editor project.

Figure 31: Administer Questionnaires. Previous (upper) and new design (lower).

Figure 32: Administer Questions. Previous (upper) and new design (lower).

Figure 33: Gameplay experiment activity sequence.

Figure 34: Hot Cherry Casino Game Machine.

Figure 35: Tool to summarize emotions by average and peak-end-rule.

Figure 36: Rated Face in the program FaceRate.

Figure 37: New identified face with id 2.
Figure 38: TextParser converting EZK-data for Excel. ................................................ 107
Figure 39: TextParser converting VAS-data for Excel. ................................................ 108
Figure 40: Correlation between SHORE and the average observer. ......................... 119
Figure 41: Compared correlation between SHORE and human Observers ............... 121
Figure 42: Similarity among all observers for recognizing the emotion angry. .......... 121
Figure 43: Similarity among all observers for recognizing the emotion happy ......... 122
Figure 44: Similarity among all observers for recognizing the emotion sad .......... 122
Figure 45: Similarity among all observers for recognizing the emotion surprised ...... 123
Figure 46: Paper-based questionnaire. ....................................................................... 147
1 Introduction and Purpose

There have been previous attempts to automate tedious work through the utilization of machines. During today’s ‘Digital Century’, where computers open up numerous new possibilities, even simple tasks rely heavily on machines. It is unclear whether or not these continued attempts, to reduce complexity by more complexity [1], will continue to succeed in the future. However, this does not prevent scientists and technicians from continuing to try and ‘simplify’ the world with the invention of new and complex machines.

Emotion recognition is part of everyday life. By looking another person in the face you are able to recognize and distinguish the emotions a person feels through their facial mannerisms and how you should react in a social interaction with that person. Emotional recognition serves as an important source of information for direct communication.

Computer-based emotion recognition works in a similar manner. After identifying a person’s facial features, an algorithm uses acquired knowledge to translate the facial features into the most probable emotion. The possible use of emotion recognition software offers advantages in many different professions and trades; anywhere where opinion or communication plays an important role.

Systems that can identify faces by using biometric features already exist and are successfully used in airports [2] and at major sporting events. By identifying potential sources of violence, they help preventing situations of social disturbance or tragedy. However, this only works for those who have already been identified by the police as petty criminals or felons, and therefore their biometric features have been stored within a
database. If there is a conflict between first-time offenders, it cannot be detected by biometric features. Since emotions are constantly expressed in the face, a subconscious and often unintentional occurrence, it would be possible to identify potential dangers from previously inconspicuous persons by using emotion recognition.

Another use would be in communication where the emotional constitution affirmation or contradiction could be transferred during a conference to give feedback; determining if all points within a situation were well discussed and if everything was understood by the individuals that were present. The speakers could then decide whether or not they would have to explain sub-points more precisely, contributing to a better understanding. Finally, a customers’ satisfaction with new products could be measured and affirmed using this software.

To make use of emotional recognition, a video signal, and software is all that is required. Video signals can be obtained with the use of any digital camera, and the software for the emotion recognition already exists in several different variations. New variants are constantly appearing, offering advances with the accuracy of recognition and/or a shorter execution time of the process. However, it continues to be very unclear whether or not software exists that is accurate. Usually emotion recognition software is trained through the use of exaggerated emotions [3], because the accuracy of identifying intense emotions is higher, leading to less influence by noise. Therefore studies, which are using a learning database filled with intense and false emotions reveal very little about a real applicability in everyday life. It would be interesting to evaluate emotion recognition software with real emotions. Therefore, this thesis deals with the research question if there is a difference between the recognition of emotions with computer-
based emotional recognition and human recognized (informed) emotions. If this is the case, it would be an advantage for the usefulness of computer-based emotion recognition in everyday life.

In this thesis the emotion recognition software SHORE [4] from Fraunhofer [5] (see 4.2.1.3) is evaluated on real usability in everyday life, because it can recognize emotions in real-time and is already a prebuilt model (no training required). For this purpose feelings are induced in test persons with the help of a casino slot machine, and then evaluated with the software and a test group. In addition, a questionnaire as well as an already proven digital questionnaire called ‘Moodmeter’ [6] (see 4.2.3) from the Sport University Cologne will be tested for the determination of emotions.

The SHORE software detects the four basic emotions (compare 2.1.1) angry, happy, sad, and surprised. Therefore these basic emotions are used for investigation in this thesis.

Since the expression and recognition of each emotion may differ, the following **Hypotheses** are deduced:

Hypothesis 1: *There is no difference in the recognition of the emotion angry between computer-based emotional recognition and human-based emotion recognition.*

Hypothesis 2: *There is no difference in the recognition of the emotion happy between computer-based emotional recognition and human-based emotion recognition.*

Hypothesis 3: *There is no difference in the recognition of the emotion sad between computer-based emotional recognition and human-based emotion recognition.*

Hypothesis 4: *There is no difference in the recognition of the emotion surprised between computer-based emotional recognition and human-based emotion recognition.*
The expected result is not a 100% match. Even humans are not infallible when it comes to correctly identifying emotions based on facial expressions. Usually the verbal communication, tone of voice, and previous events are incorporated in the determination of the emotional state. As a matter of fact, the computer-based determination of emotions in this thesis is reduced to a pure processing of static images of facial expressions.

However, the evaluation should answer the question under what conditions an application in daily life should be possible, and how accurate each particular emotion can be recognized, compared to typical human recognition.

The questions from the questionnaires (just described above) are planned to be used inside game machines, since errors happening when reading handwritten answers and manual transferring the answers to a database would be avoided. Therefore, a questionnaire editor should be implemented. This can aid the questionnaire designers by reusing previous questionnaires from others, not to forget typical questions, and already translated questions could be reused again for another questionnaire. There exist multiple well-founded recommendations¹ of how to create questions and what mistakes should be avoided. To automatize the check for mistakes which can be checked automatically, would also lead to a better understanding of the questions and therefore reduce the amount of mistakes. For example, incorrect spelling, punctuation or grammar may confuse the reader, making it harder to understand the questions, causing the

¹ e.g. “Recommendations for writing educational and psychological test items”, which gives 22 “recommended rules for writing test items” [76 p. 219]
subject not to give the most appropriate answer. In addition to grammatical errors, logic errors caused by incorrect questions or by overlapping answers can result in inappropriate answers. The definition of error is “an act, assertion, or belief that unintentionally deviates from what is correct, right, or true” [7]. Even a meandering question can be considered an error if the length has a negative impact on the correct result.

Since the game elements are always structured the same, containing the same inner states, questions could even be configured to be triggered by specific events. For security reasons (questionnaire and results are confidential) the XML-file can be saved encrypted using the Advanced Encryption Standard [8].

Figure 1 shows the actors that interact with the editor. The actor “Tester” is not in direct connection with the editor, but answers the questions shown on the Game Machine. The Game Machine attains its questions from an XML-File, which can only be read by the Game Machine. The editor reads and writes the XML-file.

Figure 1: Actors and connection with Game Machines
Interviewers, and possibly translators, will use the editor to allow the definition of new questions and connections to events within the game. Those events should trigger the pop up of a question. In this way, the subject can be asked directly about a feature after he has just experienced the feature within the game play.

By using a defined question catalogue (a single database) translations of questions will be easier to manage/administer/organize, interviewers can then benefit more from each other and the error-rate in questions will be reduced. Questions that are checked by many eyes can be globally corrected.

In order for the overview to not be lost due to too many questions; it should be possible to sort them into categories and defining typical questionnaires for different basic game types. It also reduces the required effort to create a complete questionnaire. This can be implemented by adding a tag-structure (a hierarchical taxonomy, similar to tags in YouTube, Drupal [9] or any blogs) and structuring all tags into a hierarchy (grouping and sub-tags in the form of tree-structures) for easy selection. Although these examples are applications from everyday life, they are created with much scientific knowledge and define with their tag-structure one of the Web 2.0 best practice [10] examples.

The questionnaire editor is a success if the functionality to create questionnaires works. An evaluation about the questionnaire editor is positive by fixing all negative experienced bugs/features, all assertions in the test programs are successful, and the code coverage has an acceptable percentage. Code coverage is a value of how much percent of the code is covered with tests.
Successful creation of the XML-file will be verified by reopening it. This will be done with Unit Testing, which are automatized tests to find more errors and to assure that refactoring will not change the behaviour of code [11] [12] [13 p. 81]. By applying the operations Create, Read, Update and Delete (CRUD) [14] [15 p. 381] [16] and verifying that the information is as defined. As much as possible the program behaviour is also validated by unit tests.

The plans and implementations to integrate a questionnaire in the casino slot machine itself were not used in the study itself, due to a change in hardware. So, albeit the questionnaire editor is working, a traditional paper-based questionnaire was used.

It was planned to use **additional biometric information** heart rate and skin conductance (sweating) with a device called Nexus. During the Game Developer Conference [17] 2011, the company Valve showed studies using skin conductance in order to measure arousal, shown in Figure 2.

![Current Hardware Solution](image-url)

*Figure 2: Valve's arousal measurement by observing skin conductance, from [18]*

The results looked quite promising, but are mainly based on the emotion fear, or a mixture of body reactions to danger/stress [19] and the skin feature to have a better grip
by adding moisture. An inquiry of skin conductance and heart rate did not show a connection to the observed emotions. Skin conductance randomly changed or increased constantly, leading to no connection to the emotions. The heart rate seemed to be rather connected to the total amount of movement. It was expected to see a higher heart rate during times of stronger arousal, but the movement of legs, turning of the body, so the use of larger muscles were the only found connection to the heart rate, making changes of the heart rate unimportant for the four observed emotions. However, the experiences made when measuring biometric features during the inquiry were summarized as report to ease up the next use of the NeXus device.

In Chapter 2 the theoretical background about emotions and their recognition is explained. Chapter 3 summarizes the works/reports made in preparation for this work. The experimental setups are described in Chapter 4, as well as the analysis methods, and finally Chapter 5 presents the research findings and conclusions are made in Chapter 6.
2 Theoretical Background

In this chapter the keystones of computer-based emotional recognition are presented, which are emotion itself and machine learning, thereupon the typical structure(s) of emotion recognition are shown.

2.1 Emotion

There are many definitions attempting to accurately define what an emotion is. Whenever it is discovered that another field of knowledge is connected with emotional states, a new definition appears, or old ones were adapted to be valid for today’s state of the art. Unfortunately, new details did not ease the definition, since a more abstract general definition is the usual method to include more cases.

An emotion is “any short-term evaluative, affective, intentional, psychological state” [20]. The word “affect” can be explained with “emotion or subjectively experienced feeling” [21]. The word emotion has its source from Latin “e-movere” (moving out/away) [20] and seems to aid humans to decide how to behave in the current environment thus helping them to survive. It has a social component (to communicate mood feelings to other individuals, even to some animals) and a motivational component (to encourage decisions), which explains why it is connected to ‘affect’; affect has its source from Latin “affectus” (influenced).

“Affects can be identified through immediate facial reactions that people have to a stimulus, typically well before they could process any real response to that stimulus.”
This is the key why emotion recognition is possible by observing facial expressions.

Emotions have their roots in evolution, since animals also have emotions. ‘The knowledge of the animals is a prerequisite for self-knowledge of man’ [23] (translated citate from Bernhard Grzimek²). In animals the preparations for ‘fight or flight’ [24] became an automated mechanism that was probably the first occurrence of what was later named emotion. After recognizing its communication potential, evolution took care to train and read those preparations and signalling vital information to like-minded others.

To measure emotions with scientific precision poses a problem, because the meaning of the word ‘emotion’ has not one all-embracing definition. However, that might change in the future, since the research on emotions is a very young field, becoming significantly more active within the last 50 years.

The following subsections deal with emotions in more detail. It should be noted that the topic has a remarkable breadth; thus only the most important and relevant aspects are taken into closer consideration instead of presenting a comprehensive and detailed overview.

2.1.1 Emotion-Research in the Past

The ‘Four Humors’ theory is probably the oldest accepted theory which attempted to explain the source of feelings in humans [25] [26]. It is based on the idea that the body has four liquids, the four elements [27], which define the inner mood. However, it is not

² Bernhard Grzimek was zoologist, book author, and animal conservationist [153].
clear where the theory about the four elements started\(^3\). The characteristic traits for the Four Humors are choleric, melancholic, phlegmatic and sanguine. The theory carried so much weight that even the ingredients for cooking were based on it. For example fresh fish is cold and wet, and therefore it should be cooked with spices (dry) in open fire (hot), in order to balance the food.

Based on Humorism [28], the theory of the four humors, Hippocrates [29] created the medical theory ‘Four Temperaments’ [30], at about 400 BC. As shown in Figure 3, the Four Temperaments are exactly what was previously defined as the characteristic traits for the Four Humors.

![Figure 3: The Four Temperaments, from [31]](image)

At about 450 AC (more than 800 years later) **Galen** summarized the medical knowledge of his time [31] [32] and wrote the dissertation ‘De Temperamentis’. The

\(^3\) There are written records [151] ascribed from the Egypt god Thoth [152] (lived about 1400 BC, was also named Hermes Trismegistus) which already deal with the four elements. And in Indian Buddhism exists an even older theory about the seven chakras with four of those seven chakras match the four elements.
work deals with the balance between the four liquids, resulting in nine temperaments: the four different temperaments each with one dominating liquid, four different temperaments with two neighbouring liquids dominating their complementary [28], and the 'balanced' temperament.

In 1827, more than 1400 years later Darwin, among others, was inspired by Bell's [33] 'Essays on the Anatomy of Expression in Painting' (1806) and 'Essays on The Anatomy and Philosophy of Expressions' (1824) and wrote 'The Expression of the Emotions in Man and Animals' "demonstrating the animal sources of human emotional life" [34]. Darwin did not share Bell's opinion that facial muscles only exist for the expression of emotions, but agreed with him that the muscles of respiration have also expressive functions. One of his illustrations on how emotions are expressed can be seen in Figure 4, which shows photographs depicted in the book 'The Expression of the Emotions in Man and Animals' [35].
Darwin’s main target was to show the origins of humans (more evidence for evolution theory), which behaviours humans have adapted from their ancestors (what can be retrieved in animals), and how emotions “are genetically determined and derive from purposeful animal actions” [34] (why those emotional movements were useful for animals). He also worked on the cultural differences between emotional expressions but mainly to find human common emotions. Unfortunately to ascribe most human traits to animals as consubstantial was not a very welcome understanding of the world at that time [36] [37]. For a whole century the book was rarely cited. Only a few studies of emotions were published during that time.

In 1884 William James published the article ‘What is an Emotion?’ [38], questioning if ones self-perception of his emotions matches with the understanding of what emotions
are. Up to that time the reactions of a human in a dangerous situation were understood in the following order:

1. Seeing danger → 2. Becoming afraid → 3. (automatically) choosing the correct reaction

William saw the automatic reaction as the first action and emotions as our understandings from actions and vital information:

1. Seeing danger → 2. (automatically) choosing the correct reaction → 3. Becoming afraid

“It is not that we see a bear, fear it, and run. We see a bear and run, consequently we fear the bear. Our mind's perception of the higher adrenaline level, heartbeat, etc., is the emotion.” [39].

In 1962 Schachter and Singer studied the creation of emotions by manipulating the vital information with epinephrine (adrenaline). “If a subject has a state of physiological arousal, with no explanation he will label this state due to the cognitions available to him. Therefore, this means that by manipulating the cognitions available to them, his or her feelings will be manipulated.” [40].

Silvan Tomkins published his ‘Affect theory’ in his book ‘Affect Imaginary Consciousness’ (1972) where he describes nine basic affects and has the opinion that those are the only nine existing ones: Joy, Excitement, Surprise, Anger, Disgust, Dissmell, Distress, Fear and Shame.

In the following years the amount of research about emotions increased more and more. Different models were created that should illustrate how emotions are structured among other parts of the body and consciousness. Emotions were separated into groups described as ‘basic emotions’ or ‘emotions one is born with’ and ‘culturally learned emotions’; however, there still exist the same problems:
Precise definition of mood, emotion, and affect

Inducing targeted emotions in a test person

Precise measurement/description of experienced emotions

One of the best-known contemporary emotion researchers is Paul Ekman. He shares the opinion that a certain number of basic emotions exist and that these can be clearly differentiated from each other, genetically conditioned equal among all humans. However, there are many different theories with different assumed basic emotions. After analyzing people from five countries (Chile, Argentina, Brazil, Japan and USA) Ekman hardened his assumption by examining more than 30km of film footage recorded by the neurologists Carleton Gajdusek [41]. The film footage contains different scattered tribes from the isolated highlands of Papua New Guinea, which Gajdusek recorded for several years. In addition, he studied infants and people who were blind from birth.

According to Ekman, the basic emotions are the ones shown in Figure 5. He only found a difference between the recognition of fear and surprise, which may pose the question if those are really two different emotions, or if our body does not make a difference between them. Maybe it is just our judgement which divides this feeling into two different words. However, since only the emotions angry, happy, sad, and surprised are observed here, the problem to classify between fear and surprise is not relevant.

In 1978, Ekman developed the FACS (Facial Action Coding System) together with Wally Friesen to recognize emotions. Today the FACS is still used by psychologists to measure the facial expression by external observation. Also Policemen and FBI are making use of this measurement. Even a TV show ‘Lie to me’ [42] is based on Ekman’s
researches [43] [44]. Figure 5 shows the seven basic emotions from the show ‘Lie to me’, by Tim Roth [45].

Figure 5: Ekman’s seven basic emotions, expressed by Tim Roth, from [45]
2.1.2 Emotion Recognition

The standard human method for recognizing emotions is to look at the face of another person. The brain automatically adapts the facial expression by minimal activation of one's own facial muscles. This copying the facial expressions creates the same feelings inside the body.

So we copy the expression (either by expressing the same muscles in our imagination, or by constructing our own model of perception with the seen facial expressions). In either case, we are able to feel the seen emotions and understand the emotions of the other person through internalization. This way of recognition can also be seen in animals, and can be seen as a blueprint for computer-based emotion recognition.

2.1.3 Other Aspects of Emotions

Emotions are influencing a large number of different fields. This could explain why it is so difficult to find a definition that satisfies all aspects. Emotions influence our memory: Experiencing intense emotions aids to memorize the causing events and also other events that are thematically or chronologically related [46]. To test this hypothesis, participants were asked to manipulate a sentence with a neutral, negative or taboo (highly arousing) word. The number of memorized words was seen as an indicator for the ability to memorize the sentence. Words were proven to be highly arousing (producing physiological arousal) by measuring the skin-conductance response [47]. The results showed that the higher the arousal (even negative) the better the memory.

Emotions are used as communication: It is common to smile when meeting with a person, in order to show sympathy, especially in work environments. This way of communicating emotional signals is used to smooth the way for professional and social
environment. In the model ‘Affective Social Competence’ [48] sending and receiving emotions is understood as an important social skill. The model defines four skills which are built on each other: understanding that an emotional message has been sent, determining its emotional meaning, understanding the message combined with the emotional message, and finally managing/filtering the affective information. They should define how well a person can understand a sent emotion.

The perception of emotions is influenced by emotions: The accuracy of emotional recognition of others depends on one’s own emotions. This was verified by putting participants in a specific mood (neutral, happy or sad) with music [49]. Afterwards participants were asked to assign emotions to images with different facial expressions. Results showed that a sad mood produces a negative bias (a sad primed participant recognizes happy emotions with a lower accuracy) and a happy mood produces a positive bias. “When the emotion of another person is not in line with how one feels, people have difficulty recognizing the emotions of the other person.” [49].

The expression and recognition of facial emotions differs between cultures, albeit it is said that basic emotions among all civilizations are equal. Studies that delved into cross-cultural patterns [50] [51] [52] [53] in emotion recognition had difficulties to discover if cross-cultural emotional communication satisfies more an absolutist, or a relational model. The absolutist model focuses on “absolute differences in communication accuracy across groups” [54] by assigning a defined skill of expressing and perceiving emotions. Whereas the relational model focuses on the interaction between expressing culture and perceiving culture by looking for matching expressions and thus perceived emotions through similarity. The compared cultures were Indian,
Japanese and American. Due to the comparatively small sample and the variance among people of one culture to recognize emotions, the correlations were insufficient to answer the absolutist or relational model question. However, the results clearly showed that cultures differ in the accuracy to perceive emotions expressed by another culture.

**Gender-related handling of emotions** seems to be acquired by a varying frequency in the usage of emotional words during childhood. A study was carried out with 40-70 month old children [55]. During that time children seem to be imprinted to emotional communication, since they talk and behave towards different emotions in the same way before that time. The study showed that probably because of the increased use of emotional terms with daughters (especially talking about sadness), they learn to use those emotional terms more often than sons. “It appeared that parents’ initial conversations about emotion at younger ages influenced children’s later use of those words.” [56].

In social environments the interpretation of behaviour and therefore the **identification of emotions differ by gender.** Specific emotions are socially more accepted for only one particular gender. A crying baby (using a jack-in-the-box [57]) was manipulated with a name label that gives a random gender [58]. The test showed that emotions of the baby are more often understood as anger if it seems to be of male gender and more often understood as sad if it appears to be of female gender. Together with similar results when observing the behaviour of child participants [59], it seems that “Not only are children encouraged to express emotions that are stereotypical of their gender, but they also may be perceived to be expressing the stereotyped emotion for their gender.” [56] In situations, where not all information is available, we are making
use of gender-specific stereotypes to conclude from understood behaviour to gender or to conclude from understood gender to behaviour.

2.2 Data Mining

‘Data mining’ deals with the systematic search for new patterns in large data sets [60][61]. Those patterns can then either help to split the data into different groups, or reduce dimensions by finding relevant and irrelevant data (feature selection) [62]. The selected features can then be used to train a computer to recognize a pattern (machine learning) [63] and predict values.

Described as a metaphor; Data Mining finds the relevant ‘tools’ (relations), and the Machine Learning trains the computer how to use the ‘tools’ in the most efficient way.

2.3 Machine Learning

Machine learning is “a branch of artificial intelligence in which a computer generates rules underlying or based on raw data that has been fed into it” [64]. With machine learning the computer can be trained to predict an outcome. It can predict the outcome by using knowledge; to build a model out of previous information. It comes in use when an outcome should be predicted in cases where only the relevant indicators are known and a connection exists between outcome and its influencing indicators, but the connection is too complex for humans to build a formula.

The outcome is either an estimated value, or a classification, depending on a model and what should be forecasted. Age or emotion can be estimated with a number for the
age or a percentage of intensity for one emotion. However, age could consist of different categories, and the result would be category middle-age fits best, or the person is happy.

There is no machine learning algorithm that has the ability to think like a human, to understand the world and understand emotions by feeling them with a consciousness. Instead, the small ‘world’ that the computer should know must be trained. This training, the phase when the model is built, is called the learning phase. Usually the trained data consists of a number for each indicator, and the correct response (output) that is expected for the given numbers (as input).

Applications can be found in various areas, from determining the best times for maintaining the city power grid [65] to the prediction of the next election [66]. In the field of emotion recognition machine learning methods are used to find faces, identify faces [67], and even to identify the facial expressed emotions [68].

Well-known algorithms include decision tree, support vector machines, Bayesian networks, clustering analysis, principal component analysis, and artificial neural networks [69].

Machine learning algorithms differ not only by required space and runtime during generation and use of model, but also by the ability to improve the model (learn more) after completing the initial learning phase (without having to create the entire model again).

2.4 Computer Based Emotion Recognition

Emotions are mainly recognized by observing facial expressions even if hints for emotions can be basically found by observation of any human behaviour that is
influenced by emotions. The computer-based approach of emotional recognition also follows the natural way to recognize facial expressions (contraction of each facial muscle) at first. Since computers cannot 'feel' the effect that a specific combination of contracted facial muscles has, it is required to analyze the combination. The muscle contractions are used as input for a Machine Learning algorithm, which will simply output the probability for each emotion. But first, as described in Section 2.3, it is required to train a model by using data combined with its known (labelled) output. The input data could be facial expressions itself, but usually images with facial expressions are used as input and the facial expressions are extracted by utilizing Computer Vision. This method is beneficial, since an emotion recognition will later be used to process images, and should use the same method to extract facial expressions when training the model.

When processing videos, facial muscle contractions can be identified by tracking action units (the motion of facial muscles) over a period of time [70]. Since every value represents one dimension, and the face has twenty [71] mimetic muscles (muscles used for facial expressions) [72], a feature extraction [73] aids to select only important information by filtering values that only provide identical information, as it is typical for dependent variables.

To measure if a new way to recognise emotions is more effective (higher accuracy or more classifications), the accuracy is estimated by testing it with real-declared emotions and then compared with the accuracy of other methods. The common way to measure this accuracy is called ‘10-cross-fold correlation’ [74]: The training-database is divided into 10 randomly composed stacks of data. The accuracy will be measured 10 times with
the same randomly divided data by omitting a different stack each time, training with 9 stacks and testing with the remaining stack to calculate the validation set accuracy. Finally, the accuracy is calculated by averaging all 10 validation set accuracies.

2.5 Evaluation with Questionnaires

Evaluation (Latin: valere = be valid/worth; English: value)\(^4\) means to assess worth and shortcomings of a new or modified product. Evaluation methods can be divided into 3 groups [75] [76]:

1. Moderated Evaluation
2. Survey Evaluation
3. Data Mining Evaluation

A Moderated Evaluation is an interview with one person, two persons (Joint Interviewing) or more than two persons (Focus Group). It takes more time when evaluating one person at a time but this can quickly bring new ideas in motion and provides a deeper insight into opinions about products (especially causes of unpopularity). Survey Evaluation uses surveys and questionnaires to collect larger amounts of opinions, but with limited response options. Data Mining Evaluation tries to identify trends and relationships by using statistics. In this case, the data is already available and has to be only evaluated.

\(^4\) Defined as “analysis and evaluation of a situation, especially as accompanying research of an innovation. In this case, evaluation of efficiency and performance review for the purpose of checking the suitability of the export documents in trials.” (translated from [167])
Questionnaires are “obtrusive measurement procedures meaning that they require the cooperation of the respondent” [76 p. 212] and “are reflective measurement procedures, meaning that they try to measure unobserved constructs” [76 p. 211]. A questionnaire can be split into single questionnaire items; consisting of a question (a text) and a response. "Two main types of response modes can be distinguished: Free-response, where the person must create his (her) own response [...] and Choice, where the person must choose one option out of a number of given options.” [76 p. 217] “The most common item types are multiple-choice and true-false-items.” [76 p. 218] “Multiple-choice item consists of a stem and a number of options.” [76 p. 218] Asking for attitude and personality, the most common response is closed-ended (choice).

There exist many closed-ended item types, as shown in Figure 6. A response can be divided to either frequency of occurrence or endorsement of a statement. An endorsement can be answered with either All-or-None (if a statement fits to the subject or not), or an Intensity scale (bipolar scale where the endpoints have opposite meanings).

![Figure 6: Classification of closed-ended questionnaire items, from [76 p. 224]](image)

24
2.6 Recent Emotion Recognition Research

Kasiran and Yahya already discussed the challenges of collecting customer satisfaction in an automatized way [77]. It complains that previous customer evaluations are only based on the perception of the customer, but does not discuss the problem of how it can be assured and/or proven that the emotions which are identified through recognizing facial expressions are really based on the same induced emotions. They describe that the first step is identifying the best technique in recognizing facial expressions.

The University of Amsterdam created a website [78] [79] to measure the emotional effects through evaluating the facial expressions of images. It offers to automatically analyse the emotions of uploaded images and show percentage values for six emotions. Also a smartphone app was published that recognizes the emotion happy. It is designed to automatically take an image as soon as the person is smiling (and not before) [80]. The company even cooperated with Unilever for a research about ice cream taste and the number of smiles [81]. At the same time another commercial website [82] from the same University claims to be able to analyse user satisfaction through emotional recognition. However, the opinions about its accuracy are divided [83] and there is no proof of a high test accuracy to represent the accuracy of real emotions.

Previous study work entitled “Methods to detect human emotions” [84] dealt with different methods of emotion recognition. Among the analysed methods like acoustic, visual, and skin conductivity, it was found that visual recognition is the most promising
one by reaching the highest accuracy. Therefore facial emotion recognition was chosen for correlation analysis in this thesis.

However, a measured high accuracy in test condition is not a guarantee for a high accuracy in the actual recognition of emotions. No estimated accuracy was found for real emotions. Also it is still not clear if the inner feelings, which are labelled with emotion words, really represent what people display with their facial expressions.
3 Previous Work

The following sections are a collection of work done as preparation for this thesis. Section 3.1 was created as a first report, deepening the knowledge about emotions as well as their recognition. Section 3.2 is the R&D, in which the implementation of a questionnaire editor is described.

3.1 Previous BRSU Report

After explaining in 3.1.1 the different methods that can be used to get receive information about emotions, in section 3.1.2 are combinations of different measurements described, and in 0 are the results summarized for uni-modal and multimodal models. This section was made as research and development project at BRS-University.

3.1.1 Emotion recognizing methods

This section is divided into sections for relevant senses and vital information. In every section the basic appearance, recognition methods, efficiency and problems are specified. The intention is to give an easy overview about how it works and how useful it can be.

3.1.1.1 Acoustic

The voice is required as a source for this method of emotion recognition. Since the combined acoustic results of all changed elements is audible, and demultiplexing is hardly possible (maybe only for speech recognition and voice features), there exists two unknown areas. 1: the state of each voice-influencing component. 2: the classification which state is representing which emotion.
The following steps are used for acoustic features:

1. Considers every detectable feature (every single feature may contain important information).

2. Use statistical functions (mean, standard deviation, min, max and range) in order to find the most influential features for emotional information (detect as many features as possible).

3. Detects similarities between features and emotion classes and select the fewest sufficient features (for a fast processing) which play an important role for the emotions.

Features: Albeit the complex creation of voice, there exists many features which give good hints about the emotions of the speaker. Usually we can see to which gender we talk and interpret emotions in voice different, depending on male or female speaker. For emotion recognition this information is not required additionally, because the mean average of the fundamental frequency already gives this information (one feature which is already considered). In order to recognize emotions, all different kinds of possible emotion information candidates could be used, and the most informative information then be selected in a first phase.

- Pitch (fundamental frequency F0) [85] [86]
- Formant frequencies [87]
- Loudness (energy) and duration [86] [88]
- Spectral Features (Voice quality and other Mel-Frequency Cepstral Coefficients) [88] [89]
- EigenFFT [90]

**Pitch:** It was discovered that emotions are mainly located in the higher frequencies (Emotional voices were processed with a low-pass filter with 160 Hz cut-off frequency. With that change the hit rate recognizing the correct emotion decreased).

![Figure 7: Conclude emotion by frequency with frequency range of meaningful signal, from [86]](image)

An approach was to move the cut-off frequency between 100 and 5000 Hz and compare the remaining meaningful signal (calculate the envelope of a speech signal) from the original speech signal, shown in Figure 7.

**Formant frequencies:** Every acoustical system has resonant frequencies which create a peak in an acoustic frequency spectrum [87]. Besides the influence of the pronounced vocals, the formant frequencies are influenced by the articulation, age, gender, and the physical properties of the resonating body [91], which itself is influenced by the contracted muscles, and the humidity of its surface (which is influenced by emotions). The main problem in formant frequencies is tacking and obtaining dominant candidates [92]. Those frequencies are described by F1, F2 and F3. Higher numbers are normally not relevant.
**Loudness and duration:** For example an angry person uses a much louder voice than a sad person. In order to calculate the energy, the sum of all squared frequency amplitudes is used.

**Spectral features:** As spectral features, the Mel-Frequency Cepstral Coefficients (MFCCs) are used. They are a spectrum representation in equally spaced frequency bands on the Mel scale [93].

![Figure 8: 20 calculated Eigen-FFT, from [90]](image)

**EigenFFT:** “The concept of Eigen-FFT is similar to an Eigen-face, that is, in some sense, any given high dimensional face image might be represented by a small number of parameters.” [90]. The main idea is to find a basic set of FFT-patterns. Each voice can be considered as a combination of those standard patterns [94]. Figure 8 shows an example for the calculated Eigen-FFT feature vectors. Since the representation can contain any of those vectors, this representation creates a 20 dimensional feature vector for every signal.

**Achievements:** In [89] many different speech features were compared. The results in Table 1 and Table 2 show that speaker-independent emotion recognition (no learned data from the speaker) is far away from being unerring. Speech may give additional
hints about emotions, but on this state it is only useful as additional assistant data for emotion recognition.

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Happiness</th>
<th>Fear</th>
<th>Boredom</th>
<th>Disgust</th>
<th>Sadness</th>
<th>Neutral</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>90.5%</td>
<td>4.7%</td>
<td>3.9%</td>
<td>0.0%</td>
<td>0.7%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>127</td>
</tr>
<tr>
<td>Happiness</td>
<td>7.0%</td>
<td>76.0%</td>
<td>5.6%</td>
<td>4.2%</td>
<td>5.6%</td>
<td>0.0%</td>
<td>1.4%</td>
<td>71</td>
</tr>
<tr>
<td>Fear</td>
<td>11.5%</td>
<td>7.2%</td>
<td>75.3%</td>
<td>0.0%</td>
<td>2.8%</td>
<td>1.4%</td>
<td>1.4%</td>
<td>69</td>
</tr>
<tr>
<td>Boredom</td>
<td>1.2%</td>
<td>3.7%</td>
<td>1.2%</td>
<td>80.2%</td>
<td>2.4%</td>
<td>1.2%</td>
<td>9.8%</td>
<td>81</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.0%</td>
<td>6.5%</td>
<td>6.5%</td>
<td>10.8%</td>
<td>73.9%</td>
<td>2.1%</td>
<td>0.0%</td>
<td>46</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>8.0%</td>
<td>0.0%</td>
<td>85.4%</td>
<td>6.4%</td>
<td>62</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>6.3%</td>
<td>0.0%</td>
<td>2.5%</td>
<td>91.1%</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 1: Confusion matrix for speaker-dependent emotion recognition, from [89].

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Happiness</th>
<th>Fear</th>
<th>Boredom</th>
<th>Disgust</th>
<th>Sadness</th>
<th>Neutral</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>65.8%</td>
<td>17.4%</td>
<td>14.2%</td>
<td>0.0%</td>
<td>0.7%</td>
<td>1.5%</td>
<td>0.7%</td>
<td>127</td>
</tr>
<tr>
<td>Happiness</td>
<td>26.7%</td>
<td>38.0%</td>
<td>12.6%</td>
<td>4.2%</td>
<td>15.4%</td>
<td>0.0%</td>
<td>2.8%</td>
<td>71</td>
</tr>
<tr>
<td>Fear</td>
<td>33.3%</td>
<td>5.7%</td>
<td>74.8%</td>
<td>0.0%</td>
<td>7.2%</td>
<td>1.4%</td>
<td>4.3%</td>
<td>69</td>
</tr>
<tr>
<td>Boredom</td>
<td>1.2%</td>
<td>1.2%</td>
<td>18.5%</td>
<td>59.2%</td>
<td>1.2%</td>
<td>6.1%</td>
<td>12.3%</td>
<td>81</td>
</tr>
<tr>
<td>Disgust</td>
<td>6.5%</td>
<td>34.7%</td>
<td>8.6%</td>
<td>6.5%</td>
<td>34.7%</td>
<td>4.3%</td>
<td>4.3%</td>
<td>46</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.0%</td>
<td>1.6%</td>
<td>1.6%</td>
<td>12.9%</td>
<td>3.2%</td>
<td>62.9%</td>
<td>17.7%</td>
<td>62</td>
</tr>
<tr>
<td>Neutral</td>
<td>1.2%</td>
<td>12.5%</td>
<td>5.0%</td>
<td>12.5%</td>
<td>0.0%</td>
<td>6.2%</td>
<td>62.5%</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 2: Confusion matrix for speaker-independent emotion recognition, from [89]

Problems: First of all the user must have a reason for using his voice. Voice controlled commands or communication with other humans are the most common purposes. If used in a noisy environment or a locality where multiple voices may appear at the same time, the recognition rate may swiftly decrease to a level where no information can be obtained. In order to prevent this, the speaker should be interested in a good recognition. This can be done by an interaction with voice controlled commands (if the error ration of the recognition of the commands is below the patience of the speaker).
It is also not easy to recognize disgust (even for humans) and to differentiate boredom and neutral [95].

**Improvements:** If recognition is permanently active, preprocessing may improve the performance for the case of neutral emotion (hardly any emotional expression) [96].

Until now there exists no system to recognize a person by voice and at the same time improve the emotional recognition rate by storing different feature combinations from that person.

The main lack in acoustic recognition system is the uncertainty which feature influence which emotion and what the most meaningful feature is. There have been evaluations between different features, based on neural networks, in order to recognize the most relevant features. Unfortunately the underlying data was based on voices from actors instead of recorded voiced from everyday life. It would also be interesting to see a comparison between different noisy environments, because depending on the feature, noise has a variable influence.

### 3.1.1.2 Visual

The face of a person will be used as source. The main problem for any face emotion recognition method is the number of features which influence the expressions in the face. Every feature is another dimension (otherwise even more problems occur when reducing the number of relevant features). The lack of a high number of test data (proportional to the dimensions, problem complexity) is the cause why dimension reduction must be used. Without feature reduction the processing time grows beyond computability [97].

**Features:** Tracking can give additional information about the face muscles.

- Geometric positions [98]
- Gabor wavelet coefficients (through image convolution)
- Facial movements [99] [70]

**Geometric positions:** After finding the face and normalizing its size, fiducial points (exact located points in the face which seem to contain important information about emotional expressions) will be searched [100].

![Image of fiducial points on a face](image)

**Figure 9:** Importance of each fiducial point according to sensitivity analysis. The importance is illustrated by the size of a point, from [98]

The distance and the relative sizes between different fiducial points are the extracted features. Depending on effort, accuracy of detection and the correctness of the recognition those fiducial points have different importance. As shown in Figure 9, cheeks and forehead carry only little useful information<sup>5</sup>.

---

<sup>5</sup> Actually the forehead may contain important emotion recognition information, but because of the problems in detection the right spot and the differences between every human in that spot, the hit rate for this point is hardly relevant.
**Facial movements:** Action units (AUs) are used as facial movements, selected from FACS [101]. "The AUs have some relation to facial muscular motion and were defined based on anatomical knowledge and by studying videotapes of how the face changes its appearance" [102].

![Facial motion measurements](image)

Figure 10: Facial motion measurements, from [70]

Figure 10 shows an example which AUs are mainly used and in which direction the area moves from neutral. Exploration of associated AUs showed that (Figure 11) "the bottom half of the face is almost disjoint from the top portion, except for a weak link between AU 4 and AU 11." [70]

![The learned TAN structure for the facial features](image)

Figure 11: "The learned TAN structure for the facial features", from [70]

**Gabor wavelet coefficients:** GWCs are used at the found fiducial points. Gabor wavelet can be seen as Fourier transformation (requires homogeneity to recognize a
stored pattern) of a small area by using gauss filters. “Gabor filter can be easily adjusted to get a content localization in detail both in space and frequency domains and it has multi-resolving ability and tunable focus” [103]. It performs very well as for texture classification [104], but is sensitive to illumination [103]. Similar to Eigen-faces, GWCs give similarity to stored patterns (Eigen values). In order to reduce redundancy in Eigen values, a filter bank can be used combining multiple filters [103]. Figure 12 shows the coverage of image information with 48 filters.

![Gabor wavelets and their coverage of the spatial frequency plane, from [105]](image)

**Achievements:** Even if an up to date report [103] states 100 % hit rate for disgust, happiness, neutral, sadness and surprise, a result of 100 % hit rate is either a recognition of images which have already been used as base for training data (no facial-independent emotion recognition) or the sample (12) was much too low and test images were not chosen randomly.

However, compared to other documents the following results appear to be confident: In 2002 the worst case for [106] was “Sad” with 41,26 % accuracy (Figure 13). In 2006 [74] “Fear” was detected with the worst accuracy of 76 % (Table 3), but it was probably a person-dependant test. And in 2007 [106] “Sadness” was detected with an accuracy of
88 % (Table 4) for person-dependant and 60 % (Table 5) for person-independent recognition.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Neutral</th>
<th>Happy</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>79.58%</td>
<td>1.21%</td>
<td>3.88%</td>
<td>2.71%</td>
<td>3.68%</td>
<td>5.61%</td>
<td>3.29%</td>
</tr>
<tr>
<td>Happy</td>
<td>1.06%</td>
<td>87.55%</td>
<td>0.71%</td>
<td>3.99%</td>
<td>2.21%</td>
<td>1.71%</td>
<td>2.74%</td>
</tr>
<tr>
<td>Anger</td>
<td>5.18%</td>
<td>0.19%</td>
<td>85.92%</td>
<td>4.14%</td>
<td>3.27%</td>
<td>1.17%</td>
<td>0.30%</td>
</tr>
<tr>
<td>Disgust</td>
<td>2.48%</td>
<td>1.50%</td>
<td>28.23%</td>
<td>4.63%</td>
<td>3.68%</td>
<td>7.13%</td>
<td>1.77%</td>
</tr>
<tr>
<td>Fear</td>
<td>4.66%</td>
<td>4.21%</td>
<td>2.28%</td>
<td>2.61%</td>
<td>0.70%</td>
<td>0.97%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Sad</td>
<td>13.61%</td>
<td>0.23%</td>
<td>1.85%</td>
<td>2.61%</td>
<td>0.70%</td>
<td>0.97%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Surprise</td>
<td>3.17%</td>
<td>0.80%</td>
<td>0.52%</td>
<td>2.45%</td>
<td>5.73%</td>
<td>1.08%</td>
<td>86.22%</td>
</tr>
</tbody>
</table>

Table 2: Person-dependent confusion matrix using the TAN classifier

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Neutral</th>
<th>Happy</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Sad</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>76.95%</td>
<td>0.46%</td>
<td>3.39%</td>
<td>3.78%</td>
<td>7.35%</td>
<td>6.53%</td>
<td>1.50%</td>
</tr>
<tr>
<td>Happy</td>
<td>3.21%</td>
<td>77.34%</td>
<td>2.77%</td>
<td>9.94%</td>
<td>2.75%</td>
<td>3.97%</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>14.33%</td>
<td>0.89%</td>
<td>62.98%</td>
<td>10.60%</td>
<td>1.51%</td>
<td>9.51%</td>
<td>0.14%</td>
</tr>
<tr>
<td>Disgust</td>
<td>6.63%</td>
<td>8.99%</td>
<td>7.44%</td>
<td>52.48%</td>
<td>2.20%</td>
<td>10.90%</td>
<td>11.32%</td>
</tr>
<tr>
<td>Fear</td>
<td>10.06%</td>
<td>0.53%</td>
<td>0.52%</td>
<td>73.67%</td>
<td>3.41%</td>
<td>8.77%</td>
<td></td>
</tr>
<tr>
<td>Sad</td>
<td>13.98%</td>
<td>7.93%</td>
<td>5.47%</td>
<td>10.66%</td>
<td>13.98%</td>
<td>41.26%</td>
<td>6.69%</td>
</tr>
<tr>
<td>Surprise</td>
<td>4.97%</td>
<td>6.83%</td>
<td>0.32%</td>
<td>6.41%</td>
<td>2.95%</td>
<td>4.38%</td>
<td>74.11%</td>
</tr>
</tbody>
</table>

Table 4: Person-independent average confusion matrix using the TAN classifier

<table>
<thead>
<tr>
<th>Neutral</th>
<th>Joy</th>
<th>Surprise</th>
<th>Angry</th>
<th>Fear</th>
<th>Hate</th>
<th>Sad</th>
</tr>
</thead>
<tbody>
<tr>
<td>79 %</td>
<td>87 %</td>
<td>94 %</td>
<td>83 %</td>
<td>76 %</td>
<td>86 %</td>
<td>88 %</td>
</tr>
</tbody>
</table>

Table 3: Emotion analysis test results for facial-independent emotion recognition, from [74] using tenfold cross-validation with 340 samples

---

"Database is divided randomly into ten roughly equal-sized parts, from which nine parts are used for training the classifiers and the last part is used for testing. This procedure is repeated ten times so that each part is used once as the test set" [75]

36
<table>
<thead>
<tr>
<th>70 samples</th>
<th>Anger</th>
<th>Happiness</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>75,7 %</td>
<td>4,3 %</td>
<td>10,0 %</td>
<td>8,6 %</td>
<td>1,4 %</td>
</tr>
<tr>
<td>Happiness</td>
<td>2,9 %</td>
<td>81,4 %</td>
<td>15,7 %</td>
<td>0,0 %</td>
<td>0,0 %</td>
</tr>
<tr>
<td>Neutral</td>
<td>12,9 %</td>
<td>10,0 %</td>
<td>68,6 %</td>
<td>8,6 %</td>
<td>0,0 %</td>
</tr>
<tr>
<td>Sadness</td>
<td>1,4 %</td>
<td>1,4 %</td>
<td>8,6 %</td>
<td>88,6 %</td>
<td>0,0 %</td>
</tr>
<tr>
<td>Surprise</td>
<td>1,4 %</td>
<td>1,4 %</td>
<td>2,9 %</td>
<td>0,0 %</td>
<td>94,3 %</td>
</tr>
</tbody>
</table>

Table 4: Confusion matrix for face-dependent emotion recognition, from [106]

<table>
<thead>
<tr>
<th>70 samples</th>
<th>Anger</th>
<th>Happiness</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>80,0 %</td>
<td>2,0 %</td>
<td>16,0 %</td>
<td>0,0 %</td>
<td>2,0 %</td>
</tr>
<tr>
<td>Happiness</td>
<td>6,0 %</td>
<td>76,0 %</td>
<td>18,0 %</td>
<td>0,0 %</td>
<td>0,0 %</td>
</tr>
<tr>
<td>Neutral</td>
<td>2,0 %</td>
<td>20,0 %</td>
<td>68,0 %</td>
<td>8,0 %</td>
<td>2,0 %</td>
</tr>
<tr>
<td>Sadness</td>
<td>2,0 %</td>
<td>8,0 %</td>
<td>28,0 %</td>
<td>60,0 %</td>
<td>2,0 %</td>
</tr>
<tr>
<td>Surprise</td>
<td>0,0 %</td>
<td>4,0 %</td>
<td>6,0 %</td>
<td>2,0 %</td>
<td>88,0 %</td>
</tr>
</tbody>
</table>

Table 5: Confusion matrix for face-independent emotion recognition, from [106]

**Problems:** The common computer vision problems: Camera quality, calibration, illumination, threshold value for contours, finding/tracking the face, exact localization of important points, change in illumination, multiple faces, glasses/beard/bangs on the forehead and a long processing time. In order to deal with those problems, calibrations and normalizations are applied.

Nevertheless, since AUs and GWCs can be detected very reliable and humans can simulate and recognize facial expressed emotions easily, visual emotion recognition is giving the highest hit rate. There is not a 100% correct ground-truth, as should be understood in the nature emotions and their many definitions. However, data can be collected proper and rated proper. The higher hit rate in facial expressions lies in its better purpose for more complex communication, since the communication is made by
looking into the face of a person. By using facial expressions, much more information can be transmitted than with other communication ways.

Combining different approaches, theoretically all basic emotions should be well recognized. Happiness and Surprise are well recognized while sad has a very low recognition [70].

Some reports have in common that fear will be misunderstood as surprise, anger and happiness [103] [102]. This seems to reflects the same misunderstandings which human themselves do. The difficulty to express and recognize fear is a source of that problem [98].

**Improvements:** Neural networks\(^7\) or Hidden Markov Models\(^8\) are required for learning the connection between features and emotions. As mentioned before a dimension reduction is required.

For this a nonlinear (Isomap, Local Linear Embedding und Laplacian Eigenmaps), geodesic distance measurement, should be preferred instead of a linear (PCA\(^9\), LDA\(^10\))

---

\(^7\) Artificial neuronal networks are some kind of simple simulation of a human brain, so also a learning phase is required. Things which cannot be easily defined can be automatically recognized as long as there is a correlation between features and things to recognize. It is a common application for early warning system, time courses analysis, image processing, pattern matching etc.

\(^8\) HMMs are a version of Markov Models (a statistical model, comparable with a finite-state machine that has a random probability for the use of each state transition and each state has its own probability to different outputs). Hidden Markov Model means that the inner states as well as the current state are not observable, except for the known outputs. They serve for the purpose to calculate the most probably current state, depending on the output information [164]. HMMs are also used for pattern matching and speech recognition.

\(^9\) Principal Component Analysis is the conversion of a set of possibly correlated variables into a set of values of linear uncorrelated variables [165]. It can be done by the generation of eigenvectors from the calculation of all covariances. This method is also used as lossy image compression. Elements with lower influence can be ignored with almost the same image information. [160]

\(^10\) Linear Discriminant Analysis is a partitioning an area into different class-areas. It can be compared with casting a shadow. Depending from which angle the light comes from, in the worst case an image may
dimension reduction, because linear reductions do not reflect the extracted information (they measure the Euclidean distances between features). The reduced dimensions representing facial expressions (hidden units in HMMs) should be defined between five and seven dimensions. Below that threshold too much information is lost, above gives no new information. In order to improve performance, 256x256 pixel sized faces are sufficient (since emotions are mainly detected through the low spatial domain, 16x16 is the absolute minimum) [98].

By using statistical analysis and evaluating the recognition rate, the most influencing features can be detected. This is done by using the Monte Carlo algorithm “Cross-entropy method” [107], using TAN11 or just by removing a feature and comparing the hit ratios.

The dimensional reduction with Isomap appears to be sufficient with three dimensions, because the variance has an inflexion (break-point), see Figure 14.

![Figure 14: Trough inflexion three dimensions are enough, from [102]](image-url)

---

1. Tree-Augmented-Naïve Bayes is an algorithm for learning the dependencies between the facial features. It also uses state transitions with probabilities, is not fixed to left-to-right orientation, but also follows the rule not creating a circle between the states [161].
The mapping varies from person to person, but within cloud spaces. Isomap can help in normalizing different faces in order to put the same individual emotions in the same areas, as Figure 15 shows. For more information about dimension reduction, I recommend an interesting blog: [109]

3.1.1.3 Other sources

In addition to search for emotions in elements that are mainly used for communication, the body also shows changes in other areas. Unfortunately, special hardware or actions are required for those methods. Nonetheless, those ways offer high detection rate for emotions that are hardly detectable with common methods [110]. For example surprise and fear always influence the pulse.

**Features:** Keyboard strokes with pressure sensors [111]

- Gooseflesh
- Vital Information [112]
- Behaviour
Since emotions and have different influences on the human face, voice and behaviour, the detection of emotions can improved by combining different methods [113].

**Keyboard strokes:** Appropriate hardware for measurement is required and may be expensive. It is based on different pressing of buttons when boredom or angry. Force, duration and series button pressing can be extracted as features. It is insufficient as single source for emotion recognition, but the “accuracy in recognizing anger and happiness can be improved significantly if “ face recognition is “assisted by keyboard-stroke information” [114]

**Gooseflesh:** Chills (strong emotions) can produce gooseflesh. Illuminating from a low angle and observing the results with an optical sensor gives good results [115]. Those chills are based on basic biologic signal systems of primates. For example if a mother and child are losing the sight, this can cause chills in the second growth when hearing the call again. For humans it is still unclear if music exists which produce chills in every human [116]. Since the change of gooseflesh gives too rarely hints about basic emotions and only observable by an unhandy sensor, it is neither necessary, not improving recognition results.

**Vital Information:** By using biosensors for electromyogram, electrocardiogram, skin conductivity and respiration changes, it is possible to detect positive/negative high/low arousal. The recognition for person-independent vital information is about 75% [117]. Interesting is that the physiological characteristics of emotions cannot be cheated as easy as facial mannerisms. On one hand it requires a fixed connection to specific sensors and may be refused for everyday-life because of the invasion of privacy. On the other hand it may give an automatic system for collecting emotion recognition data for speech/face
emotional recognition with very low wrong identifications for person-dependant emotions.

**Behaviour:** In order to detect the viewing rates and reaction when watching TV, to normal facial emotion recognition was added information about the control of the TV programs. Switching to another channel may show boredom or anger while not changing may show happiness. But much more features were extracted: The rate and interval of channel switching, volume adjustment and recall button pressing. Also start time watching TV and stay period will be processed [118]. No detailed recognition results are available about main recognition or supporting recognition system.

**Achievements:** Until now all additional modal emotion recognitions seem to be in their infancy. Keyboard-stroke and vital information may give interesting results in the next years, but due to the additional work and expense for sensors and for being under examination, basic emotions will be recognizes only with optical and acoustic features while the medicine may advance in analyzing the vital information in more detail.

### 3.1.2 Combining specific features

Since no simple rule for reading/showing emotions exist, a learning system is commonly used. Hidden Markov Models [95] [74] or neural networks will be used as way to read emotions, as already mentioned in the visual section. The positive effect of those learning systems is the independency of the specific used features. Any feature can be used to train a model. Visual, acoustic and even features from vital information can be combined into one model for emotion recognition. By adding features, any additional input-type has an error ratio and may ruin the whole recognition.
However, any report about fusing different input-types shows improvements in recognition. In 2007 [106] the detection of “Sadness” was improved to 68 % (Table 6) and in 2009 [119] “Surprise” was improved to 42 % (Table 7).

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Happiness</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>[106]</td>
<td>80 -&gt; 82 %</td>
<td>76 -&gt; 76 %</td>
<td>68 -&gt; 70 %</td>
<td>60 -&gt; 68 %</td>
<td>88 -&gt; 92 %</td>
</tr>
</tbody>
</table>

Table 6: Change of visual emotion recognition when adding acoustic features, from [106]

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Happiness</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>[119]</td>
<td>55 -&gt; 100 %</td>
<td>89 -&gt; 89 %</td>
<td>69 -&gt; 73 %</td>
<td>90 -&gt; 90 %</td>
<td>20 -&gt; 42 %</td>
</tr>
</tbody>
</table>

Table 7: Change of the acoustic recognition when adding visual features, from [119]

All emotions have virtually always an improvement in recognition when using a multimodal recognition. This confirms that methods and their accuracy to detect an emotion vary among the specific emotion. At least some acoustic features must be perpendicular to visual features in order to improve the accuracy. In [119] is said that “‘anger’ and ‘happiness’ were better recognized from the audio modality” and “‘disgust’, ‘sad’ and ‘surprised’ emotions” were better recognized from the visual modality.

3.1.3 Conclusion

Although visual recognition of emotion is difficult due to finding and tracking objects, it seems to be the most promising unimodal emotional recognition. Comparing the analyzed emotions with the categorized emotions in drawing books, there seems to be a higher potential.

\[12\] Values may differ in 1-5% hence bad visible exact values.
Figure 16: Different expressions of happiness: Disingenuous, accretive, caress and doleful, from [120]

Categorizing different kinds of each basic emotion (see Figure 16) may help to separate them with a much higher accuracy. For example if one smile can be recognized as disingenuous smile, it may be understood why some facial expressions are weighted high for more than one basic emotion.

Since humans have multiple emotions at the same time, the best approach might be a feature extraction [73], classification and weighting for different emotions at the same time. There is already work which explored all additional problems when working with continuous emotions [102].

Since different approaches may deliver different results for individual emotions, I suggest it should be examined how the values from different approaches can be combined in order to improve the overall recognition rate. In addition, to analyze if the cultural background influences the different recognition rates (as implied in sets with different cultural faces and their results).

Since most studies are based on simulated emotions (professional actors emphasize the emotions in their learned way) there exists a differentiation between arbitrary and involuntary shown emotions [121 S. 10]. Using acted emotions as basis for emotional recognition may cause a higher error rate on real-world data [122]. In order to solve that discrepancy, a larger data set for artificial emotion expression, and natural emotion
expression with consistency among different cultural backgrounds may be useful. Monte Carlo algorithms, Fast Fourier transformations and probability models are developed since many years and gain more and more influence. Examining the results over the past years, applications should be able to read the value of different basic emotions with an average accuracy of 90 % and above.

3.2 Preliminary Research and Development Work

The following subsections describe the creation and verification of a questionnaire editor. The main focus lies in a dynamic approach that allows adding more functionality by additional programming, as well as a good structure to keep an overview when the content grows to an unexpected size.

But first, since the game elements are displayed using the interface controls wxWidgets [123], based on a custom XML-structure. Therefore a short introduction to wxWidgets is given.

3.2.1 wxWidgets

"wxWidgets is a C++ library that lets developers create applications for Windows, OS X, Linux and UNIX on 32-bit and 64-bit architectures as well as several mobile platforms" [124]. Although the library offers much more possibilities, it is mainly used as a control library by using a custom defined structure in XML-files. The structure of controls is not strictly defined (since their interpretation is not implemented); it should conform to the existing structure for controls on game screens. Structure is shown in Figure 17.
Inside the tag "<Widgets>" any number of controls can be placed. Controls used in the company were ImageControl, Button and Label. All controls have the same possible events. Since ImageControl has no text property and Button requires multiple images just to display its different button states, Label is chosen as text displaying control for questions. Figure 18 shows the XML structure of a Label control. In order to use the control, the parameters listed in Table 8 will be adjusted.
### Table 8: Parameter in the Label control

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>xpos, ypos, zpos</td>
<td>Position of the control; Zero is the upper left corner</td>
</tr>
<tr>
<td>width, height</td>
<td>Size of the control</td>
</tr>
<tr>
<td>Text</td>
<td>The displaying text of the control</td>
</tr>
<tr>
<td>Justification, YJustification</td>
<td>Horizontal and vertical alignment of the displayed text</td>
</tr>
<tr>
<td>FontEx</td>
<td>Type and size of the font</td>
</tr>
</tbody>
</table>

#### 3.2.2 Implementation

This section is structured by grouping classes, which are dealing with the same problem and or topic.

**Serialization:** “Serialization is the process of converting a data structure or object into a format that can be stored [...] and ‘resurrected’ later in the same or another computer environment.” [125]. Within this project, serialization is a transformation between the real objects and their String representation.
All loading and saving commands are using the same custom serialization implementation except for the XML-File export. Every group of data will be serialized individually.

<table>
<thead>
<tr>
<th>Group</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>Manage Game Events (states in Game Machine)</td>
</tr>
<tr>
<td>Question</td>
<td>Manage Questions</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Manage Questionnaires</td>
</tr>
<tr>
<td>Tag</td>
<td>Categories Hierarchy</td>
</tr>
<tr>
<td>User</td>
<td>User Management (Login, Id, Password)</td>
</tr>
</tbody>
</table>

Table 9: Grouped classes that are serialized with a custom serialization

Two classes (Figure 19) are responsible for the serialization of each group (all groups are listed in Table 9). The first class is called data and contains all the values which should be saved and loaded. The second class is called management and offers access to the first class. Management is a singleton, meaning that it can be referenced by multiple classes but only exists as one instance. Whenever it is accessed for the first time, the instance is created and filled with data from its connected file. If the instance is destroyed, all its data will be written onto its local file. Instead of directly overwriting the old data, it will rename the old data for backup use and delete older data. If for some reason loading is not successful, the backup data will be loaded. Only if both loading attempts are unsuccessful, the new instance will be created from scratch (without any previous data).

The standard Serialization was avoided as much as possible, due to the constant overhead required for the serialization of common objects [126] [127] [128]. The custom Serialization was initially based on another custom implementation [129] from the website “The Code Project” [130] and has been completely rewritten a few times. A
class called DataAccessor offers easy use for serialization. Subcategories grant access to 
the path of the executable file, the ApplicationData folder or any custom defined folder. 
Every subcategory has a property named data. In order to serialize an object, it is 
assigned to the property data. If the property data is read, the deserialize operation will 
be executed. Both calls require a parameter for the target filename.

An optional encryption password can be used. As soon as a password string is 
provided, serialization and deserialization will use the encryption algorithm Shauna 256. 
The password string is not used directly for encryption. Instead the hash function of the 
password string is generated and used to encrypt the results of the normal serialization or 
to decrypt the file contents before applying deserialization.

![Diagram of EventManagement class]

Figure 19: The class EventManagement as example for Serialization

**Game Events:** All games are based on the same state machine. When a new game is 
created, a template game with the typical defined states will be used as a basis and then
filled with new content. By defining all games states in the editor, questions can be
triggered on a chosen state transition while no additional implementation code for the
new game itself is required. This game event representation is possible, because all
states are represented in game machines as normal strings (as long as the name of the
game events fit exactly the names of the states in the game).

Since it is unknown whether the template games states will be changed at a later date,
or whether a new game will have new states supporting new features of this special
game, it is required to be able to change the game events during runtime.

It is likely that only one game template is used the whole time. Therefore one
template can be defined as a standard template by receiving the flag ‘default’. Whenever
a new questionnaire is created, the default game events collection will be selected for the
triggering of questions in the game machine.

![EventData class diagram](image)

**Figure 20: The class EventData to manage existing game machine events**

The implementation is simple. EventData (Figure 20) contains all information.
“Events” holds all game events, ‘Groups’ holds all collections of game events.
‘NewDefaultGroup’ contains the ID of the standard collection of game events.
A game event (Figure 21, left class) is represented by a name and an id. A collection of game states (Figure 21, right class, Field “Events”) is a list of integer, which only contains the ids of the game states. The method ‘ToString’ defines the String representation of the object. The benefit when overwriting the standard ‘ToString’ method is that the object can be used directly in any standard control without additional display code (since controls commonly use the objects containing ‘ToString’ method to display them).

**Question:** To define a question as an object is a bit more complex. Observing its roles shows that the following behaviours, shown in Figure 22, need to be implemented:
1. The user must be able to edit the question.

2. In order to show the user how the question would look, the question must have a method to preview its content.

3. When the question should be shown in the game machine, it must have a definition of how it has to be displayed in the game machine.

Since it should be possible to define any other question type, a plugin functionality to load any additional question types should be implemented.

The preview of a question can be based on the code that is used to display the question in the game machine. In order to avoid a screen formatting definition for every single question, a question is always part of one distinct type of question, by which its formatting is defined. Therefore a question always returns two pieces of information to define its display: one (controls and formatting) is a definition of the controls for this
question category, the other (content, fill controls with text) is the plain text for each control. For example if a questionnaire consists of 10 multiple-choice questions, 8 with three answers and 2 with four answers, then only two question categories are used. If it will be 200 multiple-choice questions, all with three answers, only one question category is used.

If a more dynamic approach will be used, it would be necessary to define a script-like language in the XML file, which allows the dynamic creation of controls and dynamic positioning. Unfortunately, when defining additional question types, this would require the code to be extended in the game machine as well. To use multiple definitions for multiple different versions of one question type is a workaround to still have a high flexibility without additional required coding.

**Questionnaire**: Questions are grouped in a so-called questionnaire. Since it should be possible to add question groups to a questionnaire, a questionnaire can contain multiple questionnaires itself.
As seen on Figure 23, the list of questions and questionnaires only contain the IDs of the elements. Questionnaires contain two different artefacts about the tags (named ‘Tag’ and ‘Tags’ in the class ‘QuestionnaireInfo’). ‘Tag’ is all tags directly assigned to the questionnaire, while ‘Tags’ contains the counted tags from every question which is included in the questionnaire. If a questionnaire contains 3 Questions and each Question has the Tag “XYZ”, then the questionnaire has stored for Tag “XYZ” the number 3.

This requires new counting on every change of a questionnaire, except for just using questionnaires or displaying them.

The custom function ‘ToStrings’ serves a more detailed representation of the questionnaire. When selecting a questionnaire out of hundreds of others, the name may or may not give enough information about its content. Since a questionnaire can contain
hundreds of questionnaires/questions, the normal ‘ToString’ method cannot give a good overview. Instead, ‘ToStrings’ lists all contained questionnaires and all questions. It is used as Tooltip\textsuperscript{13} in TreeView and Listbox controls, to give a better impression of what is represented by a particular questionnaire.

**Tags:** Since the best shape for the hierarchy is defined by the users itself, a structure with Tags was chosen to allow any kind of hierarchy. A Tag is a category that the user can use to filter questions through categories. Questions are filterable since any of those categories (even multiple) can be assigned to a given question, as shown in Figure 24.

![Figure 24: Assigning and filtering with Categories](image)

**Figure 24:** Assigning and filtering with Categories
Left form displays the assigned Categories of the selected Question in the right form.

\textsuperscript{13} Tooltips are commonly used elements in GUIs. They appear when a mouse is hovered over a control with activated tooltip and display detailed information of how to use the control or what is represented by the control. Since Tooltips are only displayed when their information may be of use, they help to keep a program look simple while explaining elements in detail without overwhelming the user with too much information.
When a tag is created, it will be classified either as group or category, depending on the position it is placed within a tree-structure. A group is a collection of categories and cannot be assigned to questions. But categories, the items that are contained in groups, can be assigned to questions. Every group automatically has the tag ‘Other’ which filters for all questions, which are not assigned to any normal category in that group. This way all unassigned questions can be easily found. As soon as a question is assigned to at least one category of a group, it is not assigned to the tag ‘Other’ anymore.

By allowing multiple groups and free definable nested categories (subcategories), any kind of structure can then be defined, as long as it can be defined with a tree structure.

![Figure 25: Classes to offer a defineable hierarchy](image)

The tag hierarchy is build up with the list TagOrder (Figure 25) which contains all TagGenre and the order which the user has chosen. Every TagGenre is represented only by its ID. The Object can be looked up in the Dictionary TagGenres. Finally the list, tags,
contains every tag. Its property parent refers either to a genre or to another tag. A positive value represents a tag, while a negative value represents a genre. The class TagGenreInfo has additional variables in order to define a special behaviour. AllowMultiple allows the user to assign multiple Tags from one Genre to one question. MustAssign requires that the user must assign at least one Tag to a question. The compliance of the special behaviour is not enforced. Instead, a violation is marked by a red background color.

User Management: The user management (Figure 26) is very minimal. No user rights management is implemented, it only serves the target to be able to track which user created and changed an item.

Figure 26: Classes for the user management

The class is serialized encrypted. “Username” is not case-sensitive and an auto completion feature should aid finding the right username. Certainly there is no auto complete feature for the password.
Export To XML-File: In order to export an XML-File, the custom definition of wxWidgets (section 3.2.1) has to be used. This is done by storing a template for a screen definition (Figure 27) and a template for a label control (Figure 28) as resource in the application.

Figure 27: Excerpt from the stored resource XMLBase
String replace commands are used to change the according values, automatized by a class that returns the complete screen definition.

To be compatible with any other standard software, a standard Serialization is used for the XML-file export. Since there are problems to Serialize IDictionary implementing generics (see Serialization above), the use of Dictionaries are avoided by using simple Lists containing a custom Class that has only the generic parameters Key and Value. Different screen definitions and the text for all questions is stored in lists of a class XMLFile (Figure 29).
Figure 29: Classes for exporting a questionnaire to an XML-file

All Dictionaries are marked with the “XmlIgnoreAttribute()”, prohibiting a serialization to avoid errors with IDictionary. The method “Process” is called once the class is filled to transform the dictionary contents to the List of Entry. Finally a standard Serialization is initialized and its resulting string is stored in a file.

**Export to Word:** Since different designers may want different designs of their questionnaires, when exporting them to word, the design should be individual configurable. Also, since not every design is known and designs may change, it should be possible to easily change a design for a user. The designs are defined by using templates which are nothing else than small Word documents. Whenever a multiple choice question is exported, its according word template is used as design. In the Word document are fields, represented by text (e.g. #1#, #2#, #3#, etc.), which will be replaced by the according text/answer of the question. A user that wants a specific design is just
required to create a new copy of the Word design templates and edit them as wished. Font formatting will be applied automatically to the text parts, since a word replace function is used for replacing the text fields. The software changes the headings in the templates to the correct level by memorizing the heading level in which the question is inserted, and 'moving' the level of every heading inside the word template by the same number of levels.

3.2.3 Results Verification and Validation

The questionnaire editor has now been implemented with following changes: Question was implemented with interfaces, allowing additional question types, but no additional question types have yet been implemented. Only a minimal display definition for wxWidget controls was required since the game machines are not making use of the complete wxWidget library and are based on their own company-internal definition. Questionnaires can be defined for single game aspects or as a real questionnaire. Sorting questions and questionnaires was successfully implemented by a Tag-structure. Whenever questions or questionnaires are listed, they can also be filtered by the user by selecting tag categories.

The project was evaluated, in addition to manual evaluation during implementation, in two ways: Unit tests and with a questionnaire.

Unit Tests: The editor was evaluated automatically by implementing an additional test project. In this test project, test units were created. A unit test is a software benchmark whose role is to test every single method of a class by knowing all states of a class before and after the operation. A parameter called code coverage indicates how much percentage of the code is executed at least one time during the test. A code line
with an “if ... then ... else” is only completely covered if both code blocks are visited.

Code coverage of 100% is not definite proof that it is 100% error-free. Misunderstood functionality may work exactly as the programmer expected but may still be seen as bug if it is not the expected behaviour for the target users. Also the behaviour may be different if other connected objects have different content. However successful unit tests with high code coverage indicate a high probability that the code will work as expected and will have a reduced amount of errors. In the last 10 years unit testing has become increasingly popular. Unit testing helps to reduce the amount of errors in programming code substantially [131]. Taking the cost of bug fixing, dependent on the development stage of the product into consideration, unit testing is one of the more useful ‘modern’ programming methods [132]. The overall code coverage value in Visual Studio was not as predicative as expected. Whenever unit testing reaches the border of graphical user interfaces or other external objects which require a special interaction, it is much more complicated to evaluate the code. In order to deal with this, model-view-controller is one approach, where structure and changes are kept relative simple and the view does not require much code coverage since it just shows the model. In Visual Studio all classes (also forms) are completely included in the overall code coverage listing (Figure 30), giving an overall coverage of about 7%.
<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>Not Covered (Blocks)</th>
<th>Not Covered (%)</th>
<th>Covered (Blocks)</th>
<th>Covered (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClassManagement</td>
<td>9</td>
<td>45.00%</td>
<td>11</td>
<td>55.00%</td>
</tr>
<tr>
<td>Deadline</td>
<td>14</td>
<td>12.37%</td>
<td>12</td>
<td>87.63%</td>
</tr>
<tr>
<td>EventData</td>
<td>0</td>
<td>0.00%</td>
<td>4</td>
<td>100.00%</td>
</tr>
<tr>
<td>EventGroup</td>
<td>28</td>
<td>66.67%</td>
<td>14</td>
<td>33.33%</td>
</tr>
<tr>
<td>EventInfo</td>
<td>2</td>
<td>8.00%</td>
<td>23</td>
<td>92.00%</td>
</tr>
<tr>
<td>EventManagement</td>
<td>37</td>
<td>36.77%</td>
<td>65</td>
<td>63.23%</td>
</tr>
<tr>
<td>Extensions</td>
<td>182</td>
<td>100.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>GeneralSettings</td>
<td>28</td>
<td>82.35%</td>
<td>6</td>
<td>17.65%</td>
</tr>
<tr>
<td>MultipleChoiceData</td>
<td>56</td>
<td>100.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>MultipleChoiceQuestion</td>
<td>102</td>
<td>100.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>MultipleChoiceTextField</td>
<td>24</td>
<td>100.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>MultipleChoiceUserControl</td>
<td>442</td>
<td>100.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>QuestionData</td>
<td>0</td>
<td>0.00%</td>
<td>2</td>
<td>100.00%</td>
</tr>
<tr>
<td>QuestionInfo</td>
<td>0</td>
<td>0.00%</td>
<td>9</td>
<td>100.00%</td>
</tr>
<tr>
<td>QuestionManagement</td>
<td>23</td>
<td>26.14%</td>
<td>65</td>
<td>73.86%</td>
</tr>
<tr>
<td>QuestionMethods</td>
<td>345</td>
<td>83.13%</td>
<td>70</td>
<td>16.87%</td>
</tr>
<tr>
<td>QuestionTypes</td>
<td>7</td>
<td>19.44%</td>
<td>29</td>
<td>80.56%</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>0</td>
<td>100.00%</td>
<td>2</td>
<td>100.00%</td>
</tr>
<tr>
<td>QuestionnaireData</td>
<td>0</td>
<td>0.00%</td>
<td>2</td>
<td>100.00%</td>
</tr>
<tr>
<td>QuestionnaireManagement</td>
<td>21</td>
<td>24.14%</td>
<td>66</td>
<td>75.86%</td>
</tr>
<tr>
<td>QuestionnaireMethods</td>
<td>222</td>
<td>87.66%</td>
<td>33</td>
<td>12.34%</td>
</tr>
<tr>
<td>TagData</td>
<td>0</td>
<td>0.00%</td>
<td>4</td>
<td>100.00%</td>
</tr>
<tr>
<td>TagGeneralInfo</td>
<td>7</td>
<td>87.50%</td>
<td>1</td>
<td>12.50%</td>
</tr>
<tr>
<td>TagInfo</td>
<td>11</td>
<td>50.00%</td>
<td>11</td>
<td>50.00%</td>
</tr>
<tr>
<td>TagManagement</td>
<td>78</td>
<td>48.15%</td>
<td>84</td>
<td>51.85%</td>
</tr>
<tr>
<td>TriggerOptions</td>
<td>0</td>
<td>0.00%</td>
<td>11</td>
<td>100.00%</td>
</tr>
<tr>
<td>UserData</td>
<td>1</td>
<td>4.35%</td>
<td>22</td>
<td>95.65%</td>
</tr>
<tr>
<td>Users</td>
<td>0</td>
<td>0.00%</td>
<td>4</td>
<td>100.00%</td>
</tr>
<tr>
<td>UserManagement</td>
<td>39</td>
<td>27.46%</td>
<td>103</td>
<td>72.54%</td>
</tr>
<tr>
<td>XMLTool</td>
<td>97</td>
<td>100.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>XMLTag</td>
<td>4</td>
<td>7.84%</td>
<td>47</td>
<td>92.16%</td>
</tr>
</tbody>
</table>

Figure 30: Excerpt from Visual Studios code coverage of the editor project

From all custom classes, excluding user interfaces, 482 code blocks of total 2330 were covered, giving 20.69% code coverage. The modules MultipleChoiceUserControl, QuestionMethods and QuestionnaireMethods mainly deal with manipulation and updating of user interface elements. Having a low coverage on user interface defining classes as well as classes whose main purpose is to update user interfaces was accepted, since the required effort for high code coverage for those classes is much higher than the possible outcome.

**Questionnaire:** Interviewers (mainly Game Designers) were invited for the volunteer completion of a questionnaire about usability of the questionnaire editor. It is included in
this document in the appendix. Some questions were changed during evaluation in order to immediately fix some minor problems.

In total 20 people were asked to fill out the questionnaire. Three abstained and 10 did not hand in any results (or not in time). The combined results from the remaining 7 people are listed below. Their job specialization is shown in Table 10.

<table>
<thead>
<tr>
<th>Amount of people</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Game Designer</td>
</tr>
<tr>
<td>1</td>
<td>Product Manager</td>
</tr>
<tr>
<td>1</td>
<td>Game Development Manager</td>
</tr>
<tr>
<td>1</td>
<td>Sound Designer</td>
</tr>
<tr>
<td>1</td>
<td>Software Developer</td>
</tr>
<tr>
<td>1</td>
<td>Technical Program Manager</td>
</tr>
</tbody>
</table>

Table 10: Title of asked people that were asked to evaluate the editor

All questions with marks are shown with a complete statistic about the number of people that chose which mark and the resulting average mark. All comments are listed at the beginning of each section. In Section 3 was the word “Tag” replaced with “Category”. Section 8 was added later, resulting in only one received priority order.

1 **Interface**

- The action “Changing to advanced GUI” requires too many clicks!
- It takes some time to get used to it.
- Navigation is quite intuitive however for more ‘timid’ users I would consider adding some ‘Tool Tips’. For example, a Tool Tip for the Save button might read, “Save selection and Go Back to previous screen”
• Option to lock groups to user that created? (Edit / Delete)

1.1 The Look & Feel of the Interface is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

1.2 To find the elements/area I was looking for (Navigation) is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

2 Questions

• I’m looking forward to having support for other types of questions. (We learn a lot when players try to explain something to us)

• Default to second tab instead of first is confusing

• Only multiple choice available at the moment.

2.1 To define questions is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

2.2 Updating and deleting questions is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
</tbody>
</table>
3 Categories

- I am not sure what tags are, possibly because I missed the meeting.
- Change the name from “Tags” to “Categories”
- This is a great feature! Intuitive structure makes it very usable for everyone.
- Tag rename
- Sometimes the filtering system doesn’t seem to respond as I might expect to mouse clicks. It seems to be doing a multi-select with each additional click. Not a problem, really just took me a moment to get the hang of it.

3.1 It was made clear what Categories are and how they work…

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F  E  D  C  B  A</td>
<td></td>
</tr>
<tr>
<td>1  1  2  1  2</td>
<td></td>
</tr>
</tbody>
</table>

3.2 To define a new structure with Categories is…

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F  E  D  C  B  A</td>
<td></td>
</tr>
<tr>
<td>1  1  4  1</td>
<td></td>
</tr>
</tbody>
</table>

3.3 To filter questions by Categories is…

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F  E  D  C  B  A</td>
<td></td>
</tr>
</tbody>
</table>
4 Questionnaires

- Use a easier “open and save as” functionality
- Maybe option to warn if trying to remove a question from inner questionnaire
- Useful to see user who created the questionnaire

4.1 To create a new questionnaire is…

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2 Define a questionnaire in order to use it in other questionnaires is…

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5 Triggering

- You should pull the trigger to the question and not vice versa
- Some confusion with “Allow Cascade” button use

5.1 Connect a question with a game event is…
5.2 To trigger one Question by an answer of another question is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6 Export a Questionnaire

- Export to Word does not appear to be implemented (yet!)

6.1 Exporting to Word is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Sure</td>
<td>F</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

6.2 Exporting to an XML-File is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Sure</td>
<td>F</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
7 Misc

- This project is making fantastic progress! It is easy to use and it is clear how useful the final product will be.
- I sort of expect the finished product will come with a small User Manual
- We can use our license for controls from www.componentart.com
- Synchronize with database doesn’t appear to be implemented (yet).

7.1 Synchronize with the Database is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Sure</td>
<td>F E D C B A</td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

8 To define a priority for reworks/missing features

<table>
<thead>
<tr>
<th>Feature Implementation</th>
<th>Very Important</th>
<th>Important</th>
<th>Little Important</th>
<th>Unimportant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Export to Word</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synchronize with Database</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add more QuestionTypes</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rework window/control Categories</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game Events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find Question</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit Question</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls for MultipleChoice</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The user interface was already partly improved after collecting feedback by single visits with the interviewers. The inner structure of a program seems to be the least important for its users. First priority seems to be an understandable and easy to use interface, followed by secondly new ideas that are very easy to understand. Usually new ideas are not easy for others to understand, so ratings within the group differed in this part.

**Corrections:** Followed the first results of the questionnaire, the “Manage Questionnaire”-form was simplified by splitting it into two forms (Figure 31): one to search for questionnaires and the other one to edit the chosen questionnaire.
Figure 31: Administer Questionnaires. Previous (upper) and new design (lower)

The “Administer Questions”-form was also improved in the same way (Figure 32). Additionally the editing was changed to be directly applied to a clone of the question in order to immediately show how the updated version looks without changing the original question.
It was interesting to see that a more complex structure (questionnaire) with more values to change was rated less complex than a more simple structure (question) just because the questionnaire does not show all its changeable values on one window.

The idea with tags was seen as a very positive characteristic when it was explained and understood with enough detail. However, the freedom, to change the whole structure of tags led to different opinions about its usability. Tags were renamed to Group and Category, noticeably increasing understanding and rating for tags.

3.2.4 Conclusion and Future Work

The project is planned as the beginning of a larger project to also display the questions in the game machine, reading and transferring results into a database and to use emotional recognition as additional information. For this there is likely going to be more changes to the questionnaire editor as well as many new features. Least of all for the reason that the programs have to fit to each other. However, with completion of the structure the hardest part of the work is already completed. Emotional recognition will
be complex, but it can be built on the base of elaborations along with the submission of scientific publications.

Especially in the early phase of this project, while no display of the XML file in the game machine exists, the editor should still be useful for interviewers. Since until now questionnaires were mainly created with Microsoft Word, an export function to word would be very useful. There are, however, additional problems (implementation requirements) for question types.

Until now there were no explanations about how to fill out questionnaires, since interviewers support game subjects directly. However, on the way of automating everything in this project, it should also be possible to activate display information about how to fill out the questionnaire for every occurring question type. Additionally there is certainly a need to have design definitions about how every question type should look in Word.

Since player behaviour and event tracking are planned to be stored in the same large database, more changes will be required in this editor. Therefore the connection with the database and the versioning system is not fully implemented at this time. Main use and testing were performed with its offline capability. The database will be added to in the future by first allowing the user to manually synchronize online, later automatic synchronization on every start and exit of the application is planned.

Also the user management will be changed in the future. Plans to condition the editing of question elements by the owner (creator) of the element will require adding more user tracking in objects, perhaps even the functionality to transfer ownership of a question element.
Finishing this project will just start the development of additional projects and more changes of the current project state since the target usage requires interaction with code which is until now only planned.

Also there exists no application experience in the real field. Maybe there will be many new problems and feature requests as soon as it is used as a standard tool for editing questionnaires.

However, the positive results in usability show that the project was accepted by the users and was therefore a success. A first construction of a questionnaire editor was built and covers most of the commonly used questionnaire elements. The project will still require more time to implement an easier user interface, but fortunately all additional work can be extensions on top of the created inner structure.
4 Materials and Methods

Since emotion recognition should be non-intrusive for casino games, only audio and visual information may be of interest. Since typical casino game machines do not require an acoustic communication, the emotion recognition is limited to an optical facial emotion recognition.

In order to do a cross-correlation between computer recognized emotions and informed emotions, the emotion recognition software SHORE [4] from Fraunhofer IIC [5] is used for this research, because it offers a real-time recognition and has a pre-trained model. To be consistent between SHORE information and ground truth, the same basic emotions are selected: angry, happy, sad, and surprised.

In the experiment, subjects are asked to play a casino game while their facial expressions are recorded, evaluated by the SHORE software, and later compared against informed emotions. Game machine events like win or lose are used to trigger an emotion to be recognized by the SHORE software. Questionnaires were asked to be filled-out before (bias) and during the game session. The game was interrupted after five, ten, and finally after fifteen minutes. This should assure that the subjects could correctly answer questions about their current emotional status (self-informed emotions).
Since a clear ground truth does not exist for emotions (see 2.1.1), it is necessary to use a second experiment (ground-truth setup) to estimate the ground truth. To be consistent with the naming of participating experiment subjects, the subjects for the ground-truth setup are named “observers”. In the ground-truth setup, selected images, recorded from the gameplay setup, are shown, while the observers provide feedback about their judgement of the player emotions.

Albeit the underlying application of computer-based emotional recognition in a casino conditions requires online emotion recognition in real-time, offline processing is used here. This is because quality is more important than velocity for the experiment, and an offline processing enables a reconfiguration of the recognition parameters. However, in order to still have a valid setup, the emotion recognition software is used in its normal real-time processing mode with the offline data and therefore the results reflect the software’s usability for a real-time purpose.

The following sections deal with the gameplay setup and the evaluation strategy. Chapter ‘Instrumentation’ describes and motivated which sensors and scales are used. Chapter ‘Sampling’ explains when the research is valid and what approach is chosen to achieve validity, chapter ‘Data Collection’ explains how the instruments are used to collect the data, and chapter ‘Data Analysis’ lists all additional steps to make use of the recorded data and how the data is used to find results.

4.1 Description of Experimental Setup

The experiment setup consists of two phases. The "gameplay setup" (first phase) provides images, which are processed by emotion recognition software. The "ground-
truth setup" (second phase) evaluates selected images from the test setup by human emotional recognition and therefore constructs the assumed ground-truth.

In the gameplay setup, players are individually asked to play a casino slot machine while their facial expressions are recorded. The casino slot machine is shown in Figure 34. Each player is asked to play three times a five-minute game session. At the very beginning and after each session, players are asked to answer two digital questionnaires implemented in a PDA [133] called ‘Moodmeter’, as well as a paper-based questionnaire, targeting to receive self-informed emotions. An operation chart aids in the correct preparation of all devices. The flowchart that is used to follow the correct schedule for each subject is shown in Figure 33.
When taking a closer look at the gameplay setup, it will mainly be held as an observation (non-intruding, non-controlling) experiment and can be divided into four phases: preparation phase, test introduction, the real test session and finally the finishing
step to collect the recorded data and prepare the instruments for the next test session.

The required time per test session is between 35 and 45 minutes.

1. First of all, during the preparation phase, money will be inserted into the game machine. It is required since the casino game machine has a time lock which only enables a slow transfer between inserted money and money available for gambling.

2. Secondly, the participant is allowed in and is asked to fill out a questionnaire about general information (gender, age group). Then the test procedure is explained. After starting the recording of the videos, the live video preview is checked to see if the face is clearly visible, large/small enough, and the current game machine score can be read.

3. Thirdly, the subject plays the game three times, 5 minutes each game. If possible, the subject should not be distracted during that time. They can change specific game settings concerning how much money they risks per round. Between each 5 minutes, the game is manually stopped and the participant is asked to fill out the questionnaires in the device 'Moodmeter' as well as a paper-based questionnaire. Then the game is continued.

4. Finally, after 15 minutes of total play time, the participant is asked to fill out the same digital and paper-based questionnaires again. The video recordings are stopped; current money in the game machine is compared with the starting money. Moreover, the subject is allowed to pick the appropriate amount of candy bags.

5. After saying goodbye to the subject, money in the game machine is refilled, and game machine buttons disinfected. Then the setup is ready for the next subject.
Casino games are beneficial to induce emotions for the following reasons: Aiding to record a video of the face by requiring only little movement during play sessions; and greater emotions due to social/economic value of money. This is even valid if the subjects only imagine playing with their own money, even though the emotions might be lower. Nevertheless, a big win is still joyful, as well as a losing streak is enraging. Since it was not affordable to payout each subject with the won money in the end, a formula was defined to replace the won money with packages of candies. For each five Euro won (rounded up to the next full five Euro), the subject was allowed to select one additional candy package. This should give subjects a motivation to play further and to risk more money, in order to win more.

Afterwards, the recorded videos are analyzed to identify the facial emotions by a custom implementation that uses the SHORE library. Instead of processing the videos in real-time during the gameplay setup, an offline processing is used, which allows a more detailed processing of the images. Thus, it should judge the facial expressions with more precision. This also allows adjusting some parameters in the emotional recognition software afterwards and processing all videos again.
In the ground-truth setup, observers are asked to look at a sample set of facial expression images and estimate the intensity of each emotion. The images are selected from the recorded videos in the gameplay setup. For each 5-minute video, four images, rated by the software as showing emotions with the highest intensity, as well as four random images were selected.

The application is rather intuitive and there is little room for user errors. Nevertheless, the subjects are provided with a brief introduction and are asked if there are any obscurities. The control is straightforward.

Because emotions appear in facial expressions for just a short period and therefore a face usually has a neutral expression, it was not sufficient to select images randomly.
Therefore, the images were selected in two ways. Half of the images were selected by choosing for each emotion from every person the image that was rated as highest arousal for the specific emotion (according to SHORE) expression. The other half was selected by randomly selecting images (to assure that emotions may be found that SHORE may have overseen).

4.2 Instrumentation

Following subsections explain each used instrument and previews the structure of the expected data from the appropriate instrument.

4.2.1 Computer-Based Image Processing

The conversion of the recorded videos to a chart of emotions is divided into three steps:

- First, the pictures of each video in the correct time interval need to be extracted. Therefore, video editing software is used (4.2.1.1).

- Second, the processing of the pictures for emotion recognition requires a different input image format. Therefore a conversion to the file-format PGM (Portable Graymap Graphic) with an image converter is used (4.2.1.2).

- Third, the emotions are extracted from the PGM images by using a custom implementation of the SHORE library (4.2.1.3), and a CSV-file containing emotion values for the found faces in every processed image is created.

4.2.1.1 Extraction of Images - Video Processing

The recorded videos start before the gameplay session started, therefore the videos need to be cut to the exact time of the test session. To extract every single image, the
4.2.1.2 Conversion of Image Format - Image Processing

The conversion of an accepted image file format is required, since the extraction of images from video sequences only offers common formats. In order to convert the format, the tool XnView\textsuperscript{15} is used.

4.2.1.3 Emotion Extraction - Face Detection and Emotion Recognition

To extract emotions out of the images, the C++ library SHORE (Sophisticated High-speed Object Recognition Engine \cite{shore}) by Fraunhofer IIS \cite{fraunhofer} \textsuperscript{14} is used. Its benefit is to offer a ready-to-use trained model as well as the recognition of the emotions in real-time. The software processes images in the file-format PGM and detects the four basic emotions angry, happy, sad and surprised.

Its engine is based on “Modified Census Transformation” \cite{census} \cite{modified_census} algorithms. The trained model for its facial emotion recognition is loaded during first instantiation of the engine class. Neither training data, nor processing time for training the model is required, but since the model consists of 90 Megabytes of source code, about 15 MB compiled, the first initialization still does require some time loading. Once it is finished

---

\textsuperscript{14} “VirtualDub is a video capture and video processing utility for Microsoft Windows written by Avery Lee” \cite{virtualdub}.

\textsuperscript{15} “XnView is a cross-platform image viewer used for viewing, converting, organizing and editing graphical & video files.” \cite{xnview}. It can read quite a large amount of different graphic formats and allows converting images to most of the formats as well (also to PGM). The same developers also released a console application ‘nConvert’, which can automate image processings/conversions defined by scripts.
loading, any number of images can be processed with the engine. The parameters shown in Table 11 must be set when initializing the engine [137].

Remarks on general emotion recognition: To recognize the four emotions angry, happy, sad, and surprised, the parameters `analyzeAngry`, `analyzeHappy`, `analyzeSad` and `analyzeSurprised` must be set to ‘True’. According to the documentation, engine debug information, and results from comparing the detection, the accuracy is higher if `searchEyes` is activated. However, the parameters `searchNose`, `analyzeEyes`, `analyzeMouth`, `analyzeGender` and `analyzeAge` do not change the emotion recognition accuracy at all. This was verified by processing the same image sequence with different activated options and first comparing the results by building a graph, later by simply calculating the md5 hash, which also confirmed that the results are absolutely identical. Parameter `minFaceScore` was left as-is, because the face was clearly detected with the default value and only little phantom faces appeared. Since higher values lead to more false-negative detected faces, while lower value to more false-positive detected faces, a change would risk the correct detection of the face, or to much more appearing phantom faces.
## Table 11: Parameters for the SHORE Engine, names and descriptions summarized from [137]

The reason for not activating to analyse the other parameters is explained on last paragraph!

<table>
<thead>
<tr>
<th>Name</th>
<th>Range</th>
<th>Default</th>
<th>Used</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>timeBase</td>
<td>[0.0, 10.0]</td>
<td>0</td>
<td>1.0/15</td>
<td>Elapsed time per image</td>
</tr>
<tr>
<td>updateTimeBase</td>
<td>True/False</td>
<td>False</td>
<td>False</td>
<td>Auto estimate timeBase</td>
</tr>
<tr>
<td>threadCount</td>
<td>[1, 10]</td>
<td>1</td>
<td>4</td>
<td>Used CPU threads</td>
</tr>
<tr>
<td>modelType</td>
<td>Face.Front</td>
<td>Face.Front</td>
<td>Face.Front</td>
<td>What face detection and emotion recognizing model should be used</td>
</tr>
<tr>
<td></td>
<td>Face.Rotated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Face.Profile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ImageScale</td>
<td>[0, 3]</td>
<td>1</td>
<td>1</td>
<td>Resize image (first step)</td>
</tr>
</tbody>
</table>
| minFaceSize        | [0, 1]         | 0.0     | 100*100| $if > 1.0$: Minimum total number of pixels to accept as face  
|                    |                |         |        | $if \leq 1.0$: Minimum face size in relation to image size                  |
| minFaceScore       | [0, 60]        | 9       | 9      | Threshold to define when a face is accepted as face                        |
| idMemoryLength     | [0, 180]       | 0       | 150    | Time in seconds that a vanished face is kept in memory in order to retrieve it again |
| idMemoryType       | Spatial Recent All | Spatial | Recent | Conditions how old ids are matched with new ones.                           |
| trackFaces         | True/False     | False   | True   | Use another algorithm to keep track of a face, even when it is not recognized as face |
| phantomTrap        | Off/Delete/Mark | Off     | Mark   | Rules to deal with phantom faces                                           |
| searchEyes         | True/False     | False   | True   | Find the position of the eyes                                               |
| searchNose         | True/False     | False   | False  | Find the nose                                                               |
| analyzeEyes        | True/False     | False   | False  | Measure, how far open the eyes are                                          |
| analyzeMouth       | True/False     | False   | False  | Measure the opening of the mouth                                            |
| analyzeGender      | True/False     | False   | False  | Estimate the gender                                                         |
| analyzeAge         | True/False     | False   | False  | Estimate the age                                                            |
| analyzeHappy       | True/False     | False   | True   | Detect the intensity of the emotion happy                                  |
| analyzeSad         | True/False     | False   | True   | Detect the intensity of the emotion sad                                    |
| analyzeSurprise    | True/False     | False   | True   | Detect the intensity of the emotion surprised                               |
| analyzeAngry       | True/False     | False   | True   | Detect the intensity of the emotion angry                                  |
The engine can process faces by using three different facial detection modules as `modelType`: ‘Face.Front’, ‘Face.Rotated’, and ‘Face.Profile’. Since faces are never ideally looking to the front, the modules ‘Face.Rotated’ or ‘Face.Profile’ seem to be promising, since ‘Face.Rotated’ can find faces with “in-plane rotation (’-60’,’-45’, ’-30’, ’-15’, ’0’, ’15’, ’30’, ’45’, ’60’)” [137], and ‘Face.Profile’ can find faces with an additional a yaw rotation between -90’ and 90, and a pitch between -20 and 20. However, the modules mainly focus on detecting a face instead of evaluating its emotions. There would be a small increase of detected faces, whenever the subject is rotating his face. Since more possible face positions exist, the face detection would be lower with a higher amount of false detections. A personal check if faces are correctly identified could remedy that, but the analysis modules would still not be able to use any of the turned faces. The analysis modules are limited to frontal heading faces (roll -15 to 15, pitch 0, and yaw -45 to 45). So the additional detections either would be phantom images, or correct faces that cannot be analyzed due to too high identified values for roll, pitch, or yaw. Therefore ‘Face.Front’ was selected.

**Remarks for video processing:** `timeBase` should be defined as the time in seconds that elapses for a single image. If a video has 15 fps (frames per second), then it should be \( \frac{1.0}{frame\_rate} = \frac{1.0}{15} = 0.06 \). The parameter `updateTimeBase` can be used if the frame-rate is unknown or varies. During several observed test runs the parameter `timeBase` either converged to a lower value between 5 and 10 fps, or sometimes jumped up to a four-digit value, which is reminiscent of a not fully stable PID-controller [138]. Therefore, since the frame-rate is well-known, `updateTimeBase` should be deactivated.
If the parameter \textit{trackFaces} is enabled, then a face may not be lost and forgotten if it is not identified as a face during some images, since a second algorithm is constantly trying to track lost faces. There is no difference in emotion detection, since no emotions can be identified from a face that is only tracked, and not identified as a face, but the face will keep its id. If the tracked face is then identified as a face, after some frames it will still be identified as the same face as previously detected, reducing the work to join the recognized emotion data for every single id.

If tracking of a face fails the parameters \textit{idMemoryLength} and \textit{idMemoryType} enable to match a new face with previously seen faces by using fingerprints (one for each id). The rules for face recognition can be changed to more strict ones, because the comparison is relatively time consuming, especially when processing a video in real-time with multiple faces simultaneously. Since only one face is observed and the recognition does not happen in real-time, the values can be adjusted to a relatively high value. The maximum stored time, specified in seconds, for a previously identified face can be changed with \textit{idMemoryLength}. If an id was not seen for the defined time, then it will be discarded by deleting the stored fingerprint. The parameter \textit{idMemoryType} can delimitate possible matching faces. When assigning ‘Spatial’, then a face is only checked for a match if it is located at about the same position as the previous face of the fingerprint. Only with ‘Spatial’ the id of the lost face will be directly assigned to the new found face. This mode is promising if a face is slowly moving out of the visible area and comes back to the same position, or if a face will be covered temporarily by other objects. But it does not prevent a face to change to a higher id (the first estimated fingerprint did not fit), and later may change again to a lower id. When summarizing all
emotions of a face, it still requires to combine the emotions from multiple ids, with the only difference that less ids exist (less work to merge all correct), but multiple ‘gaps’ in one id may require filling with the data from another id. This leads to a higher amount of time required in order to merge all emotions from one face that was detected as multiple faces. Even when ‘Spatial’ reduces the total number of different detected faces by the use of fingerprints, the higher effort to merge the resulting data was the reason why ‘Spatial’ was not selected.

When assigning ‘Recent’, every face will be compared with fingerprints from previously detected ids as long as those previous ids were last seen in less than idMemoryLength specified time. If there is a match, then the face still keeps its id, but a new parameter ‘RecentId’ will be added to the face, referring to the matching id. If multiple ids match, then multiple parameters ‘RecentId’ with an appending number are added. This enables to combine the emotion recognition results from multiple ids while still being able to check each id manually and correct false-positive faces.

The third type, ‘All’ is similar to ‘Recent’. Instead of multiple ‘RecentId’ parameters, only one parameter contains the matching ids as a space separated list. The big difference is that the list should contain all previous ids that the face ever matched with (or the matched faces ever matched with). Therefore a phantom face in the background that vanishes and appears over and over again, would lead to a constantly growing list.

In order to reduce the amount of phantom faces, the parameter phantomTrap can be used. It is deactivated with ‘Off’, marks detected phantom faces with ‘Mark’ and deletes them immediately if detected as phantom faces with ‘Delete’. When using ‘Mark’, still not all phantom faces were recognized as such. But also the correct face was sometimes
marked as a phantom face. Maybe the static sitting person in front of the camera does not provide enough movement, and is therefore sometimes recognised as phantom. Since the inner algorithm to find phantom faces is unknown, it would not make any sense to use \textit{phantomTrap}, because of the false-positive detected phantom faces.

To reduce the amount of phantom faces without the drawback to remove correct faces as well, the parameter \textit{minFaceSize} can be used. If a face is detected, then its size is defined as \( \text{faceSize} = \text{faceWidth} \times \text{faceHeight} \) in pixels. So when setting a relative low limit, there are still some unfiltered phantom faces. However, the correct face will never be filtered in this study, since the minimal size is limited by the distance of the chair and the real size of the face. The parameter \textit{ImageScale} is left as-is (1.0) since a change requires readjusting the value of \textit{minFaceSize}. The lower processing time for downsized images does not play a role, and the image will not gain any additional information when enlarging it.

The parameter \textit{threadCount} should be set to the maximum number of cores of the CPU. However, with a quad-core (virtually 8 cores) CPU it was still required to execute 4-5 times the face detection at the same time, in order to use the CPU to full capacity. This leads to the conclusion that the parameter, windows, or both do not operate as expected.

General conditions for the software in order to recognize emotions [137]:

- The faces need to be at least 24x24 pixels and “the input image must be sharp for good results on small faces”. [137] And for a more precise analysis by using \textit{searchEyes}, the face should be at least 36x36 pixels.
• Faces should be heading front, the determined value for ‘roll’ must be ‘-15’, ‘0’, or ‘15’. For ‘pitch’ the value must be ‘0’ (‘-20’ or ‘20’ would be the next determined values), and ‘yaw’ must be ‘-45’, ‘0’, or ‘45’. Meaning that a face must be rolled between ‘-22.5’ and ‘22.5’, pitched between ‘-10’ and ‘10’, and yawed between ‘-67.5’ and ‘67.5’.

4.2.1.4 SHORE Library Utilization for Analysis Purpose

Since the SHORE face detection is built as library, an executable program has to be created. It can be based on an example console application to process a single image, but has to accept multiple images, extract the emotion information, write it in one file that can be used for further processing, and should give some calculated intermediate values in order to verify the correct execution of the operations.

Since an image stream is required, but only single images are accepted one after another, it was implemented to accept multiple images by accepting a folder as input. After reading all filenames from the given folder and storing them in a vector, this vector was processed systematically, the image loaded and the engine fed with it.

The emotion information is read and collected in separated lists for each found id. Each list contains elements that are a combination of an image and its appropriate detected emotions. After processing all images, the emotion information for each id is written in a separate file with a name that represents the id + ‘.dat’.

In order to filter phantom faces and connect multiple ids which represent the same face, an algorithm was implemented which is based on the test situation: To expect only one face at the same time. All additional faces, which appear at the same time, are phantoms. If the correct face is lost and a new face appears for the first time, then it is
expected to be the new correct face (which somehow was not tracked). The algorithm is described in more detail in subsection 4.2.1.3, as well as the construction of a file that should only contain all correct recognized faces. In addition, if a new face was found, the current image is marked with a frame around the face and a number representing its id and stored in the output. Since only pixels are directly accessed instead of a graphic function that enables to paint a text at a location, the numbers are drawn by using a 7-digit display structured pixel drawing.

To receive some additional feedback, the maximum detected value for each emotion was stored, as well as the number of images that contain each emotion with a value higher than zero.

4.2.2 Data Recording for Emotion Analysis

To capture facial expressions as well as states of the game machine and movements of the subjects, multiple requirements must be fulfilled (probably only solvable by recording multiple videos concurrently): Facial expressions must be recorded from frontal position and with high resolution and sufficient frame-rate, casino game machine states should be recorded (other angle) with same conditions (sufficient resolution to ‘see’ all information, sufficient frame-rate to track changes as fast as they occur, not overseeing any), and other subject’s states by keeping track of any movements of the subject.

4.2.2.1 Casino Game Machine Score

The current score of the casino game machine may give hints about the emotional state of the subject, since it reflects the winning and loosing streaks. The frame-rate requires maximum 1 fps, since the score only changes every few seconds. However,
since the recorded video should serve to find distinctive events (score changing events), a constant frame-rate and a synchronized time between events-video and facial expressions-video is crucial.

A synchronized time is achieved by recording both videos with the same computer. A constant frame-rate does not pose a problem, since the frame-rate can be lowered down to five fps and a low resolution. In addition, this gives the side effect to have more resources to processing the facial expression video.

The resolution is sufficient if the score can be read in the video, which is fulfilled by any webcam. The camera “Hercules Deluxe Webcam” is used for the named requirements, and is mounted upside down on top of the game machine, in order to allow subjects to read the displayed score while it is recorded. The score can still be read in the processed video by applying a digital filter (rotation 180 degree) in the video capturing software.

4.2.2.2 Emotion Analysis

Facial expressions should be constantly recorded during the gameplay setup. There are high demands on the camera. On one hand it should record faces in a high resolution to allow a good extraction of relevant facial features: It is required to have a face size of at least 36*36 pixels in the resulting image in order to recognize emotions with a high accuracy (see subsection 4.2.1.3) while still recording a large area around the face to keep the face in the visible recorded video, even when the person moves his face. Secondly it must provide a sufficient frame-rate (at least five, better more than ten) to capture even short-term facial expression (excluding micro expressions), and should
thirdly allow direct access to the video to create a video file which allows relatively easy
video editing and image extractions.

Since the emotional recognition model is trained for frontal positioned faces, best
results are achieved if the faces are recorded from the front, between subject and game
machine. Therefore, the camera should be small enough to fit between the subject and
game machine without limiting actions, and be attached at the front screen of the game
machine, directly below the wheels.

The webcam Logitech HD Pro C910 meets those requirements: It has a relatively
small size while still offering a good image quality [139] with a Zeiss-optic lens. The
webcam supports up to 30 fps with a resolution of 640x480 pixels. It can be connected
directly to a USB port and is accessible for any video capture software.

4.2.2.3 Other Actions and Reactions of the Subject

The subject’s movements as well as game machine states should be recorded to find
additional distinctive events, which do not only depend on the score (e.g. a bonus round
where a high win amount was almost reached). There are no limitations in size of the
hardware, but the recorded video should have a high resolution, since it should capture
the subject as well as the game machine completely (and each single light and button
pressing of the game machine should be easily observable).

The Panasonic HDC-HS900 is a real video camera and offers a Full HD recording
with real 30 fps [140]. The camera was placed about one meter away from the subject,
so that the test persons are filmed from the rear left. This enables both the display of the
game automaton, and the movements of the subject to be visible.
4.2.3 Alternative Emotion Retrieval

Two questionnaires are used to identify the physiological state (emotions) of the subjects. First, a digital questionnaire called Moodmeter, and secondly, a paper-based questionnaire that was created for this study.

4.2.3.1 Digital Questionnaire

The device Moodmeter was developed in the Institute of Psychology by Prof. Kleinert at the Sport University Cologne. Its focus is to track the physical body constitution of athletes on daily or weekly basis to early detect changes in the body mood [6]. This early feedback of the athlete’s constitution can then be used as a preliminary investigation if it is reasonable to change the training plan or even start to further investigate the causes for the change.

The Moodmeter is a proven digital questionnaire implemented in German that rates 12 athletic qualities: mood ("Stimmung"), level of recreation ("Erholtheit"), willingness for exertion ("Anstrengungsbereit"), self-confidence ("Selbstsicherheit"), state of rest ("Ausgeruhtheit"), willingness to communicate with others ("Kontaktbereit"), recognition ("Anerkennung"), peace of mind ("Innere Ruhe"), fitness ("Trainiertheit"), perceptual vigilance ("Aktiviertheit"), agility ("Beweglichkeit"), and health ("Gesundheit").

It calculates the qualities by asking how much a body-constitution-describing word fits for the person. The answers are 0%, 20%, 40%, 60%, 80% and 100%. Moreover an answer has to be selected within 5 seconds, in order to assure a spontaneous answer without thinking about memories to find a reason for it.
It was not yet used for rating emotions which may change within seconds and does not focus to detect the intensity of the four basic emotions which are observed by the emotional recognition software. Therefore a correlation has to be examined between the known emotions and the twelve rated indicators, in order to decide if the Moodmeter can be used as a reliable measuring instrument for emotions.

4.2.3.2 Paper-Based Questionnaire

Mainly as a first comparison between Moodmeter and emotions, an additional paper-based questionnaire, to ask the subject about the current intensity of their four observed emotions, has to be developed. There exist different tools to ask the user about their feelings, verbal and nonverbal. One would be to use cartoon faces with easy recognizable emotions, similar to Emocards [141] [142] and Geneva Emotion Wheel [143] [144]. The other contains normal text; a plain verbal scale shown to be as effective as Emocards [145], during feedback it could be revealed that Emocards was not common for many people, therefore it may be confusing and influence the tested person. Therefore a paper-based questionnaire will be created, which is directly asking for the four measured emotions and allows the user to mark the intensity for each emotion on a scale from neutral to maximum. The used questionnaire is shown in the appendix (Figure 46).

It is important that the paper-based questionnaire is not a validated questionnaire. It may give a clue about the emotions of the subjects, but it is not proven that subjects will really be able to answer the questions logically. Especially when emotions are represented with words, a rating of emotions may be misunderstood by another meaning, resulting in a different rating.
4.2.3.3 Compare Emotions with Emotional Events

The subjects are doing self-assessments when they answer the questionnaires provided for them; giving information about their own inner feelings. The emotion recognition tracks the momentary visually visible emotions by measuring the facial expressions of the subject.

In order to compare the answered questions (post evaluation) with the constant measurement of emotions, the measured emotions must be summarized to the same dimension. Therefore, a method is used that promises to deduce the physiological emotional state from instantaneous states, called the "peak-end rule" [146]. This has already been proven as one of the two common methods where the human subjective impression of an experience can be predicted. It is calculated by taking the peak (highest arousal) and final impression into consideration. The other method is an average over the entire experience period. For a series of many different experiences the average experience seems to deliver meaningful results [147]. Since peak-end rule provides generally better results [148], and with 'multi-episodic events' the average method has been proven to be effective, both methods will be tested for suitability. A correlation will then answer which of the two methods is more suitable for this case. Whether or not one method at all is suitable is unclear, because each gaming history is a series of numerous random events.

Indeed, there are references collected by the determination of current emotions, but it is unknown whether or not the expression of emotion and its intensity reflects the inner arousal, which influences how events are processed (if only because of the communicative and social use of emotions).
A software tool, shown in Figure 35, is used to determine the values for peak, last values, and average. Since it is not known if the last recognized emotion represents the last impression, the tool allows averaging any number of the last seconds of the recognized emotions.

### 4.2.4 Generation of Reference Ground-Truth

In order to have a ground-truth reference about emotions, what expression is generally understood as what emotion, there can only be used the evaluation of humans (since those also define the emotions). A tool should allow this by allowing humans to evaluate the emotions of faces in shown images and averaging the ratings.

The Windows application ‘FaceRate’ is a PC-tool for subjects to judge the intensity of the four basic emotions for selected images, as shown in Figure 36. It displays all images from a specific folder, one after the other. For each image, the user has to click on a labeled bar at the bottom to rate the intensity of an emotion between 100% and 0%.
The design of the rating bars was adopted from the paper-based questionnaire. The maximum rating on the leftmost side was adopted from the question design of the ‘Moodmeter’ (the subject works with already learned educated structures).

![Rated Face in the program FaceRate](image)

Figure 36: Rated Face in the program FaceRate

When the program is closed, it stores the relevant data, i.e. its path and its rating, for every rated image in a file. If the subject starts the program again, it will read the file and continue where it was stopped recently. To prevent data loss during an unexpected error, the program stores all ratings in a temporary file every 3 minutes.

After all images have been rated, the program automatically sends the results by email to a preconfigured email address. The results are attached to the email as a CSV-file (comma separated values).

### 4.3 Data Collection

During the data collection, multiple problems led to discard the collected data about one subject. The game machine stopped accepting 20-euro bills when its inner sensors
noticed a low amount of stored coins. It was not sufficient to bypass the photoelectric sensor checks by covering the transmitters, because accidental activation of the payout led to error codes. Whenever the machine tries to pay out coins, additional photoelectric sensors in the coin-payout-channel assist in noticing a failed payout (for example because of a stuck coin). Those additional sensors need to see a free area during start-up and an obstacle during every coin payout. Easiest solution was to fill the coin storage with 300 euro in one and two euro coins.

A cupboard needed to be closed during the play session. It has a mirror attached to the inner side of the door and created interference in the recorded video (reduced illumination and increased chance for phantom faces).

The face-recording camera was not fixed at the game machine during the first day. It was fixed by sticking it to the game screen itself. The recorded data from the first day were discarded.

Problems to keep the subjects face in the visible area for the camera were solved by disabling a camera zoom feature. Some camera settings reset after a pc restart, therefore the settings were inspected during every morning.

Whenever a subject was looking at the score display, he needed to lift his eyebrows a bit. This movement was not caused by surprised emotions, rather to have a better field of vision in the upper area that is physically a bit covered by the eyebrows. This strengthen the assumption to estimate the ground-truth by observers instead of connections with game events and informed emotions of the subjects themselves.

During the third time filling-out the Moodmeter, a subject said that he filled-out the Moodmeter in a wrong way before, by expecting the max value on the very right. It was
tried to cope with that issue by adapting the structure of the paper-based questionnaire to the structure of the Moodmeter. This way it was still apart from being intuitive, but the subjects do not have to acclimatize to two different systems.

For another subject it was forgotten to let the person fill-out the Moodmeter during a game-break. Questionnaire and Moodmeter data were discarded for both subjects.

One time the Moodmeter device crashed and was only useable after a hard-reset.

Multiple subjects were confused about the questions of the Moodmeter. Two international students needed an explanation for the Moodmeter question if they feel “ladiert” (damaged/battered).

The purpose of some questions about the current game state in the questionnaire showed to be successful. Multiple subjects were surprised that they lost money. It was realized when answering the question about the amount of remaining play money in the casino game machine. Therefore, it is not clear if the first five minutes can be used as correct subject influencing data. During the remaining ten minutes, the subjects were much more attentive about their current game money. It was tried to create this understanding by explanations that are more detailed, followed by a test-round, but in the recorded cases, it did not improve the subjects understanding of the game mechanisms.

Subjects reacted very different when they lost play-money during the play session. Except for a few subjects, it was more common to smile or even laugh when loosing play-money. It is not clear if this is a cultured habit when dealing with frustration or a way to personally distance from the game events. However, noticing that issue, it was not sure if game events aid as indications for the emotions of the subject.
Two subjects lost all their money faster than it was possible to insert it. Main reason for that is the time-locked refilling mechanism due to law restrictions. An inserted bill is not immediately useable for the game. Instead, the game has two different money counters: One for the inserted money and one for the money that can be used for playing. Only one Euro will be transferred to the playable money per minute.

Multiple subjects accidentally activated the payout mechanism by keeping a specific button pressed for multiple seconds. It took some more time to refill the machine, but the facial expressions were kept, since it was an unexpected event and offered a good chance to see a surprised face.

One subject was used to roll the face about 10 degrees to the side, leading to issues with the face detection. An additional video processing to rotate the video image corrected the problem.

### 4.3.1 SHORE-Based Console Application

The console application, which uses the face emotional recognition library SHORE, produces five different outputs.

1. A file `debug.dat`, containing all generated outputs from the emotion recognition engine, as shown in Table 20.
2. A file `stats.log`, giving a summary about all processed images and results, shown in Table 12.
Total pictures processed: 5121

Pictures with...
   Emotions: 2486
   All Emotions 0: 2562
   Tracking only: 68
   No Face found: 10
   More than 1 id: 31

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Occurrence</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry</td>
<td>14</td>
<td>46.2963</td>
</tr>
<tr>
<td>Happy</td>
<td>2465</td>
<td>100</td>
</tr>
<tr>
<td>Sad</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Surprised</td>
<td>25</td>
<td>86</td>
</tr>
</tbody>
</table>

Settings:
model: Face.Front
searchEyes: 1
searchNose: 0
searchMouth: 0
analyzeEyes: 0
analyzeMouth: 0
analyzeGender: 0
analyzeAge: 0
processing loops: 1

person.csv is combined out of:
1. 0
2. 1
3. 23

Table 12: Example content of the file stats.log.

3. One or multiple files ‘<number>.dat’ (usually starting with ‘0.dat’) containing all detected emotions for the face identified with this number (face ids are enumerated, 0 is assigned to the first id, 1 to the second, etc.) Its content is shown in Table 13.
Table 13: Example extract of the created emotions list.

4. One or multiple images, similar to Figure 37 which contain information about the ids found in the image:
5. The file ‘person.csv’ that contains the estimated emotions of the person. Its structure is the same as files described in point 3 (‘<number>.dat’) but the content is a combination of multiple face ids.

It happened several times that the application lost track of a face. The causes were leaning too far to the side, putting a hand in front of the face, bending down when filling-out a questionnaire, or rotating the head too far. An algorithm in the SHORE engine is configured to memorize faces and will try to recognize the face as soon as it appears again. This recognition is limited by space and implemented with a limitation by time.

Moreover, there were some problems when a face appeared at a different location than where it had previously disappeared. However, if a face is not recognized as a known face, then a new id is assigned to the face. Usually this happens only two or three
times, but in some cases (person is turning/tilting more than 15 degree (compare SHORE model settings in subsection 4.2.1.3) or moves the face partly out of the recorded area) it occurs numerous times resulting in a two-digit number of times. This problem cannot be solved by combining the emotion data of any id that was recognised as a face, because in many cases phantom faces (part of the background is recognized as a face) were found as well. Therefore an algorithm was defined that can correct more than 90% of the problematic cases by defining id 0 as always the correct face, and therefore assigning 0 to a variable `correct_face`. An additional appearing face while the `correct_face` is still visible is interpreted as a phantom face (in this experiment, there was always exactly 1 face visible). Furthermore, if the `correct_face` is not found, but another id appeared for the very first time, then the new id is understood as the `correct_face`. ‘person.csv’ contains the combination of all face ids which have been identified as the correct face.

To assure that the file ‘person.csv’ only contains the emotions of the correct recognized face and not any phantom face, multiple steps are required:

1. Open file ‘stats.log’ and scroll down to the bottom, to see the list which describes all ids that have been identified as the correct face.

2. Open the generated images and verify that the correct face has an id which is included in the list for the correct face, as well as verify that phantom faces are not included in the list.

   a. Phantom face was in the list:
i. Open the emotions listing file for this id (e.g. if id 16 is a phantom face, then open 16.dat) and memorize first and last image where emotions were found for this id.

ii. Open ‘person.csv’ and replace all images in the phantom image range with empty emotions (-1,-1,-1,-1).

b. Correct face was not in the list:

i. Open the emotions listing file for this id, memorize first and last image and copy the data.

ii. Open ‘person.csv’, mark the memorized image lines and overwrite them with the copied data from the correct face.

3. ‘Person.csv’ can be opened in Excel.

4.3.2 Moodmeter

The device stores the results on its SD-Card in ‘Moodmeter \GruppenAntworten ezk_VAS.txt’ and ‘Moodmeter\GruppenAntworten wkv_VAS.txt’. The translation for each word is shown in 4.2.3.1. Sample content can be seen in Table 14.
Table 14: Example for EZK answers.

With a text parser, as shown in Figure 38, the output is converted to a format that can be imported to Microsoft Excel.
An example of ‘GruppenAntworten wkv_VAS.txt’ is:

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Trainiertheit</th>
<th>Aktiviertheit</th>
<th>Beweglichkeit</th>
<th>Gesundheit</th>
</tr>
</thead>
<tbody>
<tr>
<td>27.04.2012</td>
<td>12:29</td>
<td>27.10</td>
<td>2.4</td>
<td>4.2</td>
<td>2.0</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>12:38</td>
<td>27.10</td>
<td>2.8</td>
<td>4.4</td>
<td>1.6</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>12:45</td>
<td>27.10</td>
<td>3.0</td>
<td>4.8</td>
<td>1.6</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>12:53</td>
<td>27.10</td>
<td>2.8</td>
<td>4.2</td>
<td>1.6</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>13:17</td>
<td>27.11</td>
<td>3.8</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>13:28</td>
<td>27.11</td>
<td>4.0</td>
<td>4.0</td>
<td>4.6</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>13:36</td>
<td>27.11</td>
<td>4.0</td>
<td>4.0</td>
<td>4.6</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>13:44</td>
<td>27.11</td>
<td>4.0</td>
<td>4.0</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Table 15: Example for VAS answers.

With the same text parser, the output is also converted for Microsoft Excel, shown in Figure 39.

4.3.3 Automation Scripts

*Exact file creation time:* Since multiple videos are recorded with the same computer, the timestamps should match with each other. However, since the video files do not contain any information about date and time, it must rely on the file property ‘Created’. The risk that arises from the property ‘Created’ is that as soon as a file is moved to
another location (and any copy of a file); the ‘Created’ data is reset to current time. Windows only displays hours and minutes, not seconds, albeit seconds are readable.

Therefore, a small file renaming tool was created, which creates a copy of an existing file in the current directory and names it to old filename + creation date and time with hours, minutes and seconds. As a result, the ‘Created’ information is not lost anymore when moving or copying the file.

Randomly select Images: It is too time consuming to manual use a random number generator in order to select images from every five-minute game session. Therefore a tool was created to automate this process. It will randomly select four images from the first third of the session, four from the second, and four from the last third of the session.

Measure the manual scales with crosses: The paper-based questionnaires contains a bar rating for each emotion. On one side it was 0%, on the other 100%. The plan to make that bar exactly 10cm long and measure the position of the cross did not work due to paper, margins and printer settings. The calculation of the correct percentage was too time-consuming for more than 400 bars. Therefore a tool was created which calculates the distance to a percentage value (depending on the defined maximum), and copies the result to the clipboard.

Batch processing: Even with small amounts of data, an automated batch processing for repetitive work makes sense. Since the tool for image conversion and the tool for emotion recognition are respectively console applications, both steps were automated using batch processing (see Table 19). The script processes every subfolder by calling the procedure `enterFolder` and passing the name of the folder as first argument. The procedure `enterFolder` checks if the folder with the grayscale images exist. If not, it tries
to create them by converting jpeg images in the folder to greyscale graphics. Finally, the
result folder is cleaned from the previous results and the emotion recognition program is
called. The processing time per five minutes of video ranges between 20 and 45 minutes,
excluding the time to mark beginning and end, and extracting the video frames as
images. Therefore, the total processing time for 20 samples with each consisting of three
times five minutes requires about 40 hours.

*Microsoft Excel charts:* After recognizing emotions, the maximum of each emotion
has to be found. But prior to that, noise should be filtered. It may happen that one face
feature is incorrectly recognized as an emotion (e.g. the subject is asking something and
the mouth position is understood as surprised). A median filter should help to filter out
recognized emotions that were seen for just a few frames. Then a search function finds
the highest value for each emotion and displays the according filename and emotion
value next to it. Since multiple highest values may exist, the function for second highest,
third highest, and so on is a bit different, limiting the search area. This is due to the fact
that the conditional cell references cannot directly search for the second occurrence of a
given value.

### 4.4 Sampling

To analyze the correlation between computer and human recognized emotions, a
sufficient number of images is required. When assuming a confidence interval of 95%,
the population to be at least 20,000 (formula varies minimal for larger populations), the
answers for each emotion to be randomly distributed (50%) and a maximum standard
error of 5%, then the sample size should consist of at least 377 facial expression images. This is calculated by using [149]

\[ n \geq \frac{z^2 \theta(1 - \theta)N}{(\Delta \theta)^2(N - 1) + z^2 \theta(1 - \theta)} \]

For an unknown unit value \( \theta \), the worst case can be assumed: \( \theta(1 - \theta) \leq 0.25 \), therefore 0.25 is used

Level of significance \( \alpha = 0.05 \) leading to \( z = 1.96 \)

Confidence interval \( \Delta \theta = 0.05 \)

Since the indicators from the Moodmeter do not fit into the four emotions, the practicability to use the Moodmeter for recognizing emotions is not clear and will just be observed by a correlation. The validity is ensured by a high grade of internal (assumption holds against null hypothesis) and external (sample can be generalized to population) validity.

The internal validity is lower if more other influences for the depending variables exist (excluding the influences from the independent variables that are controlled or tracked). The casino game machine serves the purpose to cause emotions that are of interest. Events of the game machine may indicate tendencies to specific emotions, but in order to assure correct identified emotions, multiple questionnaires are used as well as an additional judgement of the emotions by a group of observers. If an emotion is neither changing the vital information, nor influencing the behavior or thoughts of a person, then it is not an emotion since emotions are defined to “provide the affective component to motivation” [150].
*External validity* is only given if the sampling is representative and can therefore be generalized. Since it differs by culture how facial expressions are reflecting emotions [35], and also how accurate the expressions are recognized [52] [71] [72], it would be best to collect data from every culture. However, due to limited feasibilities, and limited observed emotions, and the results about common basic emotions [41], only a group of Europeans will be tested.

The subjects are not actors and therefore not well trained in the recognizable display of emotions. Additionally, no one was asked to show emotions; instead events were used to induce natural emotions. This ensures that the results are more easily comparable with the results of an application under normal conditions. Since the sample better reflects the population, it leads to a higher external validity. Also, as it is unknown how far gender influences emotion expression, to specify own emotions, and to recognise and describe emotions from others, it should be kept in mind to represent both sexes in equal amounts.

The sample of subjects does not play an important role for the external validity when comparing emotion recognition between computer and people, but in order to reach a minimal significance when comparing the questionnaires, representing self-informed feelings, with the recognized emotions from the software, representing the recognition of facial expressions during natural induced emotions, the sample should not contain less than 20 people.

The fact that all participants have a high socioeconomic status has a negative impact on the external validity (the participants have a higher status by being students, and more future perspectives). Secondly, it may cause a problem if money plays an important role for all subjects. Otherwise, emotions are less evoked by a money casino game machine.
Furthermore, the subjects might mistrust the ability to influence happenings in gambling machines.

Due to voluntary performance, an informed consent is not required. Anonymity is given by grouping results and for single results by replacing the names with pseudonyms.
5 Experimental Analysis and Findings

Three different evaluations have been considered:

1. Computer-based emotion analysis by correlation of the softwares emotion recognition results with the ground-truth (based on reference subjects, named as 'Observers'). This should evaluate if natural emotions can be recognized with the same accuracy as acted emotions when using emotion recognition software.

2. Measure applicability of the Moodmeter by correlating Moodmeter's indices to Peak-End ruled SHORE recognized emotions. It can also be compared to the self-informed emotions, but this only determines if the Moodmeter values 'could' show a connection to the emotions. This is not required for the scope of this work.

3. An attempt to measure the connection between questionnaires targeting self-informed emotions and the computer-based measured emotions. This may answer the question if an observing camera (nonintrusive) can replace a questionnaire that is targeting self-informed emotions.

In order to measure the correlation and statistical significance, the software SPSS Statistics was used. Spearman's rho correlation was used since the values are not normally distributed. In order to calculate the significance, a 2-tailed test was applied and significances marked with asterisks (* for significance, ** for strong significance). Additionally, Microsoft Excel was used to improve the legibility of each table.

Feeding the database: The paper-based questionnaire is manually transferred to a database. The remaining money in the game machine at the end of each five minute
interval is normalized among all subjects by subtracting the initial money at time 0 from
the values for time 5, 10, and 15.

To summarize the emotions interval of the computer-based emotion recognition to
one single value after each five minutes, the peak-end rule is used [146]. The ‘final’
emotional state is determined by averaging the detected emotions of the last three and
ten seconds. The peak is even harder to determine, because the causing event for an
expression is unknown for the software, and it is not clear if the detected emotion 100%
happy describes the same inner arousal as the detected emotion 100% sad or 100%
angry, or how to normalize the emotion arousals with each other. Therefore, by
following the peak-end rule, the experienced events should add an according bias to the
following averaged emotional states due to hysteresis (since the person remembers
previous events and judges his current situation and emotions by those).

It is not clear how intense such a hysteresis is and how long it lasts, therefore three
different approaches are built to represent the peak. One approach is to simply measure
the emotion with the maximum intensity if multiple emotions reached 100%, and then
the emotion which lasted at 100% for the longest time is understood as the peak. The
second approach is a simple average over the whole time, which may work if the peak
occurs during the beginning, but the accuracy will be lower, the later the peak occurred.
However, since casino games usually pay out low amounts and the events for a jackpot
were tracked and marked (and can be applied to the according data peak measurement
afterwards), the positive and negative events should only differ slightly. Therefore, a late
occurring peak, compared to the previous events, would not change much of the persons’
mood.
It is also unknown if a happy expression really describes a happy emotion or just a protection or trained behavior [151], maybe to better cope with frustrations in social environments. Therefore, a peak that vanishes as fast as it just appeared is most probably either noise, or a faked happy emotion that was used to hide the sad emotion.

To compare computer and human recognized emotions, the recognized emotions from each selected image, which was also rated by observers, are copied to a new database containing only those images. The ratings from each observer are added to the same database. A linear correlation function is then used to compare computer results with single observer results, computer results with averaged observer results, and observer results with averaged observer results. Interestingly, if the computer does not correlate that well to other observers, but some observers may also not correlate that well. Since computer-based emotion recognition is trained by data from humans, a computer cannot recognize emotions better than humans can.

_Correlation between computer and human-based emotion recognition:_ The videos of 27 subjects have been processed. The first four were removed due to a bad initial camera position for recording frontal faces. Each image of the entire video was extracted, converted, and processed with SHORE, which led to more than 400 000 files, and a space consumption of 150 GB. From those files, 490 images were selected in total. Half of the images were selected by choosing the highest emotion arousal from each person’s recordings by finding the maximum emotion values by SHORE. The other half was randomly selected among all images, while 11 of the randomly selected images were rejected, since the face was not found by SHORE, resulting in 479 images. The expressed emotions in the images were rated by ten volunteer observers, half male, half
female, and were correlated against the results from SHORE for the exact same images.
In the following figures, observers are named Obsn, with n between 1 and 10. The
observers 1 to 5 are female, 6 to 10 male.

Two ways are used to compare machine recognized with human recognized emotions.
The first section uses observing humans as ground-truth, while the second section uses
the self-information from the playing subjects as ground-truth. Concluding a third
section presents specific findings in more details and evaluates which hypothesis hold
out against the found results.

5.1 Computer-Based Emotion Analysis

The computer-based emotion analysis compares the emotions that are recognized by
software for emotion recognition with the ground-truth built by averaging the ratings of
ten observers.

In a correlation matrix, the data of each column is correlated with the data of each
row. A correlation matrix for this work is shown in Table 16. This correlation matrix
ends with ‘Observer10’. The last column is calculated by averaging previous cells.
SHORE represents the results from the computer-based emotion recognition by using
the software SHORE, the 10 observers represent the people that rated the selected 479
images. The rows ‘Average all Observers’ on the very left of the table is correlated to
SHORE by averaging the answers from all observers and then correlate the results with
the answers from SHORE. The correlation between ‘Average of all Observers’ and a
single observer is calculated by building the correlation between the single observer and
the ground-truth (observer average), consisting of the 9 remaining observers. For
instance, the correlation of ‘Average of all Observers’ with ‘Observer1’ is calculated by averaging observer 2 to observer 10 and correlating the results with observer 1.

The values on the last column are not directly based on a correlation, but based on averaging all single observer correlations (the previous ten columns). Therefore, in each “Avg. all Observer”-row, it is the average of correlation of all reference subjects to ground truth (average of reference subjects without the correlated subject).

The asterisks reflect the calculated significance. There is no significance if the number has no asterisk, significance with one asterisk and a strong significance with two asterisks. Since the last column is not a correlation, it has no asterisks (there exist no significance formula for averaging). The significance for the field SHORE correlated with SHORE was filled out in order to complete the table, and to know that the correlation was calculated correctly. No significance was calculated for it, but it should be very clear that it would be very significant in all four cells. It may help the reader to see that the column SHORE reflects the same data as the row SHORE.

The colors reflect the value range of the correlation coefficient:

- $[0.0; 0.1]$ – **yellow**
- $[0.1; 0.3]$ – **red**
- $[0.3; 0.6]$ – **blue**
- $[0.6; 1.0]$ – **green**
Table 16: Correlation matrix between computer and human recognized emotions.

To answer the question how well the computer-based emotion recognition fits to the average emotion-recognizing human, the ratings of all observers were averaged and correlated with SHORE. This was done for each of the four emotions. The results are depicted in Figure 40. The values are obtained by rounding the first column in Table 16.

- A strong correlation was found for “Happy”.
- A moderate correlation was found for “Surprised”.
- A weak correlation was found for “Angry” and “Sad”.

![Figure 40: Correlation between SHORE and the average observer.](image)
Since one person's emotion recognition does not necessarily fit to another person's, the averaged ratings of all observers\textsuperscript{16} were also correlated to each single observer and the resulting 10 correlation coefficients averaged to a single value for each emotion, shown in Figure 41. This gives answers to the other hypotheses:

- **Hypothesis 1 (angry) is rejected**, because the correlation is very significant, and only a weak correlation of 0.148 has been found.

- **Hypothesis 2 (happy) is accepted**, because the correlation is very significant, and a strong correlation of 0.688 has been found, which differs only little from the average observer correlation of 0.797.

- **Hypothesis 3 (sad) is rejected**, because the correlation is significant, and only a weak correlation of 0.1 has been found, which is much lower than the average observer correlation of 0.486.

- **Hypothesis 4 (surprised) is accepted**, because the correlation is very significant, and a moderate correlation of 0.286 has been found, which lies close to the average observer correlation of 0.382. Albeit the correlation lies in the lower area, it is comparable to the detection of humans, which also strongly differs for this emotion (observers 5 and 10 correlated less than SHORE to the averaged recognition).

\textsuperscript{16} To keep the correlation clean, the average ratings of all observers are constructed by averaging the ratings of all 9 other observers, comparable with the 10-cross-fold correlation [75].
The red bars in Figure 41 have the same source as the red bars in Figure 40. Each blue bar in Figure 41 is calculated by averaging the correlations among the ground-truth references (observers). For example, the blue bar for angry is equal to averaging the ten blue bars in Figure 42. Therefore, the correlation among all observers (blue bars) is shown in more detail in the following four figures.
In Figure 42, it is shown how the ‘averaged correlation between averaged observers and individual observers’ (the blue bars in Figure 41) is calculated for the emotion angry. In Figure 43 is the same connection shown for the emotion happy. In Figure 44 for the emotion sad, and Figure 45 for the emotion surprised respectively.

**Figure 43: Similarity among all observers for recognizing the emotion happy.**

The emotion “happy”, shown in Figure 43, can be recognized well among all observers, and SHORE also has a high similarity to the rating of each individual observer.

**Figure 44: Similarity among all observers for recognizing the emotion sad.**
The emotion "sad" in Figure 44 was found in this study to vary the most in recognition. Respectively for the computer-based emotion recognition exist the problem, to find a good intersection among different peoples detection.

![Graph showing similarity among all observers for recognizing the emotion surprised.](image)

The emotion "surprised", shown in Figure 45, was acceptably well recognized by SHORE, compared to the recognition values of the observers.

According to the selected images, SHORE found 98 images with angry, 278 with happy, 44 with sad, and 145 with surprised arousal. The observers saw 40 to 286 images angry (in average 148), 158 to 235 (average 186) happy, 77 to 437 (average 205) sad, and 94 to 539 (average 213) images surprised.

### 5.2 Computer-Based versus Self-Reported Emotions

In order to analyze another way between computer-based emotion analysis and ground-truth was taken into account by using the questionnaire Moodmeter as alternate ground-truth. The correlation in Table 17 represents the connection between Moodmeter and machine recognized emotions by SHORE. The rows represent different techniques.
to summarize a five-minute block of the emotions as one value. For each emotion, the average of the recognized intensity within the last three seconds was calculated, within the last ten seconds, and for the whole five-minute session. In order to use the peak-end rule, the maximum value was also calculated. Significance (asterisk) and color-coding is explained at Table 16.

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Moodmeter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level of concentration</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>-0.030</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>-0.005</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>-0.020</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>0.006</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>0.011</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>-0.098</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>-0.095</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>0.006</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>-0.009</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>-0.014</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>0.025</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>0.081</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>-0.079</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>-0.032</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>-0.044</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>-0.007</td>
</tr>
</tbody>
</table>

Table 17: Correlation between SHORE and Moodmeter.

Since all correlation coefficients are very low, only weak correlations were found. It was expected to find in column ‘Mood’ a minor better correlation coefficient, compared to the other columns. There are no correlation coefficients that could be accounted as medium correlations, as well as only little significances, which also do not count, since
the correlation is too low. The found weak correlations are seen as noise and no assumption can be made.

Table 18 shows the constant weak correlation between the same SHORE summarizing values and the self-reported emotions. No connection or assumption can be made since only weak correlations exist. Significance (asterisk) and color-coding is explained in Table 16.

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Angry</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>0.036</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>0.068</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>0.090</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>0.143</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>-0.075</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>-0.007</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>-0.037</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>-0.050</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>-0.182</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>-0.179</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>0.025</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>0.010</td>
</tr>
<tr>
<td>Avg 3 sec</td>
<td>0.070</td>
</tr>
<tr>
<td>Avg 10 sec</td>
<td>0.064</td>
</tr>
<tr>
<td>Avg 5 min</td>
<td>0.057</td>
</tr>
<tr>
<td>Max 5 min</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Table 18: Correlation between SHORE and self-reported emotions.
6 Conclusions and Future Work

The correlation between computer and human recognized emotions showed that the software was moderately accurate in recognizing emotions that were also easily recognized among all humans; while it had a much higher imprecision if the recognition of an emotion also differs among humans. The emotion "surprised" is the only exception, because most subjects recognised a surprised face with ease, while the library's accuracy was much lower.

An explanation may be the simple definition of a surprised face while the interpretation of such detected characteristics is not always that clear. When a person looks at an object located above the face, then the eyebrows may cover a part of the currently important field of viewing. A natural mechanism is to raise the eyebrows, appearing surprised for an observer who does not know about the reasons. Detecting the pupil locations (if resolution allows the detection) may reduce this error, but a good recognition may require even an understanding of the surrounding complex world. As long as movements cannot be connected with the surrounding world, it can only be achieved to reach a high accuracy for communicated simple emotions. Even humans among each other have problems to recognize others emotions, as the varying accuracy showed. If the reason for emotions with lower recognition accuracy is individual expressed emotions, then acted emotions may be a positive approach for training emotion recognition as well as a good base for defining a common emotion expression language among humans.
On condition that the average rating among all observers represents the basic truth about emotions, the determined correlations show that the recognition of emotions can be used for marketing if an evaluation of the emotion happy is sufficient. As per the experimental analysis results here, the accuracy is not sufficient in order to use the system for security observing systems, since the recognition rate for angry lies far below the recognition rate for humans themselves. If a market analysis that is only based on happy is reasonable and representative is not completely clear, because it is not known if recognized emotions represent the true emotions, since it was not successful to connect the computer recognized emotions with the self-reported emotions or the Moodmeter.

One possible explanation for the non-significant results between summarized happy emotions and self-informed emotions is the ambiguity for happy facial expressions [152] [151]. Another was the unclear use of the Moodmeter for multiple subjects. The Likert scale is arranged to increase from right to left, which was for multiple subjects not what they first expected, leading to a few discarded questionnaires due to misunderstandings. The chosen words in the questions were also confusing for many subjects. However, a correlation with many significant values was found between Moodmeter and the self-informed emotions, as shown in Table 21 on page 146. So there had either been not enough serious events, Peak-End rule does not help in summarizing events, or expressed emotions differ from the real experienced emotions. A higher sample may have rendered the results more significant, but more importantly, another method to summarize multiple experiences (if possible) has to be explored.

Recapitulating what has been compared, it was due to not finding any significance between the Moodmeter and the summarized emotions, since the mood is not defined by
the sum of all experienced emotions, and the Moodmeter does not target to estimate emotions, there exist many moderate correlations to emotions, but since every word correlates with any positive emotion and negatively correlates with any negative emotion, emotions cannot be differed with the Moodmeter. Only a one-dimensional mood (which is already estimated by the Moodmeter) can be compared with the total combination of emotions.

The SHORE library from Fraunhofer is closed source with all consequences. The source is protected, which means that modifications require a long way by communicating with the source-code owning company, but on the other hand it is not required to deal with the details of the emotion recognition. The comparable low accuracy (compared to the results of other emotion recognition in the last years, see section 0) is reasonable for the libraries purpose to offer a real-time emotion recognition. It may be modified for other purposes on request, and maybe other inputs will be integrated and offered for the library in the future.

Emotion Recognition is a constantly improving field of research. There could be used other ways to measure emotions (e.g. heart rate monitoring or sweating by measuring skin conductance), but not all ways are qualified to improve emotion recognition.

As described before a previous test with subjects playing a casino game and recording heart rate and skin conductance did not show any connection. The heart rate depended on the movements of the subject, whether the person is angry, happy, or just likes to move a bit every few minutes. The sweating at the fingers is connected with being nervous, since it should give more grip to hands and feet. So a higher amount of pressing
buttons can increase it, but had no connection with game events, since different game
situations do not vary at all by the frequency of button pressings.

Since the cerebral research is working with different imaging methods, a future study
could try to assign individual brain activity patterns with corresponding emotions. This
would be a huge step forward, since the difference between felt emotions, displayed
emotions and self-identified emotions is still a very unclear research area. However, it is
not an easy task, since at least all three of these types should show in the brain.
Bibliography


4244-5375-7.


Smit, Lindsay, I. *A tutorial on Principal Components Analysis.* Dunedin, New Zealand : University of Otago, 2002.


http://en.wikipedia.org/wiki/Principal_component_analysis.

Appendix A (Thesis)

Batch Processing script (run.bat) to convert all images to grayscale and extract its emotions with the emotion recognition console application:

```
SET /A _count=0
SET _outFolder=\out
SET _inFolder=\PGM
SET _jpgFolder=\JPG
for /f "Tokens=*" %%G in (\dir /A:D /B') do call :enterFolder "%%G"
echo Finished!
GOTO :eof

:enterFolder
  REM Called for every Subfolder, with foldername in %1
  SET /A _count+=1
  echo ... processing % _count % Folder: %1
  cd %1
  IF NOT EXIST "%_inFolder%" (CALL :checkJpgFolder)
  IF NOT EXIST "%_inFolder%" (CALL :noInFolder) ELSE (
  CALL :checkFolder
  call .\Demo.exe -e face -i "%_inFolder%" -o "%_outFolder%" -1 1 -s f1000000 -fps 15 -d >PGM2Emo.log
  )
  cd..
  GOTO :eof

:checkJpgFolder
  REM If no PGM-folder exist, then it is checked if it can be created with JPGs
  echo Trying to convert JPG to PGM
  IF NOT EXIST "%_jpgFolder%" (GOTO :noJpgFolder)
  md "%_inFolder%"
  ..\convert.exe -out pgm -grey 256 -o "%_inFolder%\%%" "%_jpgFolder%\*.jp*" > JPG2PGM.log 2>&1
  echo ...conversion complete
  GOTO :eof

:noJpgFolder
  echo No Folder with JPEGS ("%_jpgFolder%") exists either
  GOTO :eof

:noInFolder
  echo No "%_inFolder%" exists in %1 (no pictures will be processed)
  GOTO :eof

:checkFolder
  REM empties outFolder or creates it if it does not exist
  IF NOT EXIST "%_outFolder%" (GOTO :checkFolderCreate)
  cd "%_outFolder%"
  del /S /Q *..* >..\Cleanup.log
  for /f "Tokens=*" %%G in (\dir /B') do rd /s /q "%%G" >>..\Cleanup.log
  cd..
  GOTO :eof
```

143
Table 19: Batch processing for image conversion and calling emotion recognition

Content of the file ‘debug.dat’:

<table>
<thead>
<tr>
<th>Image 1 of 5121 was processed (.\PGM\0000.png)</th>
</tr>
</thead>
<tbody>
<tr>
<td>= Object[0]</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---Type: &quot;RightEye.Front&quot;</td>
</tr>
<tr>
<td>---Region: Left=265.4, Top=145.8, Right=310.5, Bottom=190.9</td>
</tr>
<tr>
<td>---Marker</td>
</tr>
<tr>
<td>---RightEye: X=288.0, Y=168.4</td>
</tr>
<tr>
<td>---Rating</td>
</tr>
<tr>
<td>---Score = 64.0</td>
</tr>
</tbody>
</table>

FrameRate: 14.9999924
GallerySize: 1
ImageCount: 50
ImageHeight: 480
ImageWidth: 640
PatchCount: 1

Image 2 of 5121 was processed (.\PGM\0001.pgm)

Object[0]
|---Type: "Face"
|---Region: Left=252.9, Top=130.4, Right=391.5, Bottom=269.0
|---Marker

Table 20: Example extract of the file debug.dat.
Correlations

<table>
<thead>
<tr>
<th>Mood</th>
<th>Angry</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of recreation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingness for exertion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self confidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State of rest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingness to communicate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace of mind</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trainiertheit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptual vigilance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 21: Correlation between Moodmeter and self-informed Emotions.

The blue italic text in brackets was added afterwards as translation. The personal questions are only asked in the very beginning. The questionnaire after 5, 10, and 15 minutes only asks about the 4 emotions and the money on the casino game machine.

### Fragen zur Person: (personal questions)

**Geschlecht: (sex)**

- □ Männlich (Male)
- □ Weiblich (Female)

**Alter: (age)**

- □ 18-24
- □ 25-34
- □ 35-44
- □ 45-54
- □ 55-64
- □ 65-

**Wie fühlen Sie sich momentan? (How do you feel at this moment?)**

Bitte setzen sie ein Kreuz X an entsprechender Stelle auf dem Balken. (Please make a cross at the according position)  (does not apply at all)

- 10 bedeutet, dass es sehr intensiv vorhanden ist
- 0 bedeutet, dass das Gefühl gar nicht vorhanden ist

(10: it is very intensive; 0: the emotion is not there at all)
Wie viel Geld ist momentan auf dem Spielautomaten?

(How much money is in the casino machine at this very moment?)

___ ____ __ ___ €

Figure 46: Paper-based questionnaire.
Appendix B (UNB R&D)

The questionnaire used to evaluate the usability of the editor.
1 Interface

1.1 The Look & Feel of the Interface is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

1.2 To find the elements/area I was looking for (Navigation) is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

1.3 (optional) I see the following problems with the interface:

=================================================================================================

=================================================================================================

=================================================================================================

2 Questions

2.1 To define questions is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

2.2 Updating and deleting questions is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>
2.3 (optional) I see the following problems with the questions:

3 Categories

3.1 It was made clear what Categories are and how they work...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

3.2 To define a new Category is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

3.3 To filter questions by Categories is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Note: The text “Tags” on this page was replaced by “Categories” in further evaluations.
3.4 (optional) I see the following problems with Categories:

4 Questionnaires

4.1 To create a new questionnaire is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

4.2 Define a questionnaire in order to use it in other questionnaires is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

4.3 (optional) I see the following problems with the questionnaires:

5 Triggering

5.1 Connect a question with a game event is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>
5.2 To Trigger one Question by an answer of another question is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

5.3 (optional) I see the following problems with triggers:

6 Export a Questionnaire

6.1 Exporting to Word is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

6.2 Exporting to an XML-File is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

6.3 (optional) I see the following problems with exporting:
7 Misc

7.1 Synchronize with the Database is...

<table>
<thead>
<tr>
<th>Very Hard</th>
<th>Very Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>E</td>
</tr>
</tbody>
</table>

7.2 (optional) The following should be reconsidered:

---------------------------------------------------------------

---------------------------------------------------------------

---------------------------------------------------------------

8 To define a priority for reworks/missing features

<table>
<thead>
<tr>
<th>Feature Implementation</th>
<th>Very Important</th>
<th>Important</th>
<th>Little Important</th>
<th>Unimportant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Export to Word</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synchronize with Database</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add more QuestionTypes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rework the window/control Categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game Events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find Question</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit Question</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls for MultipleChoice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Curriculum Vitae

Candidate’s full name: Nils Reichert

Universities attended: Bachelor of Computer Science, 2009

BRS-University, 53757, St. Augustin, Germany, www.h-brs.de