Dominick Leppich, Carina Bieber, Katrin Proschek, Patrick Harms* and Ulf Schubert

DUX: a dataset of user interactions and user emotions

Abstract: User experience evaluation is becoming increasingly important, and so is emotion recognition. Recognizing users' emotions based on their interactions alone would not be intrusive to users and can be implemented in many applications. This is still an area of active research and requires data containing both the user interactions and the corresponding emotions. Currently, there is no public dataset for emotion recognition from keystroke, mouse and touchscreen dynamics. We have created such a dataset for keyboard and mouse interactions through a dedicated user study and made it publicly available for other researchers. This paper examines our study design and the process of creating the dataset. We conducted the study using a test application for travel expense reports with 50 participants. We want to be able to detect predominantly negative emotions, so we added emotional triggers to our test application. However, further research is needed to determine the relationship between user interactions and emotions.

1 Introduction

In our modern, digitalized world, User Experience (UX) evaluation is a topic of increasing importance. This results from the expectation of many industry branches that customers and the average population are able to and should perform more and more tasks independently, just with the help of online services [1, 2]. UX is "[a] user's perceptions and responses that result from the use and/or anticipated use of an interactive system. Users' perceptions and responses include the users' emotions, beliefs, preferences, comfort, behaviours, and accomplishments that occur before, during and after use" [3]. Constantly investigating users' emotions during their interactions with technical devices contributes directly to improving User Interfaces (UIs) and their UX.

Detecting the emotions of users outside of laboratories or without special hardware allows accessing data from real-life usage and takes a greater number of users into consideration. Nowadays, it is possible to recognize user emotions with the help of different data sources like facial expressions or the users' heartbeat. These data sources differ in their requirements. For example, facial coding faces challenges like illumination, pose, and occlusion [4]. Research shows it is possible to detect emotions more reliably by combining different data sources [5]. The collection of Keystroke, Mouse and Touchscreen (KMT) interaction data can be done unobtrusively as an additional data source [[5], p. 1]. At the moment, however, detecting the emotions of website users is often only possible in the laboratory with the help of appropriate technology. In order to create algorithms that detect emotions exclusively on KMT data, we first need a dataset containing KMT and emotional data. With this dataset, we can search for a connection between user behavior that can be ascertained on a technical level and the users' actual emotions, moods, and arousal. The dataset can be used as training data for Machine Learning (ML) algorithms to learn these kinds of connection automatically. As such a dataset is not yet publicly available, we decided to create our own [[6], p. 13]. We conducted a study and recorded the users' interaction with the UI and additionally recorded the users' emotions. The recording of user interaction data was non-invasive to influence the users as little as possible. All this data was merged into one dataset that contains lower level keyboard and mouse interaction data together with emotional data. Our leading research question was: "How can we record and process user interaction and emotion data to create a dataset of all user interaction with their corresponding emotions at any given time?" We publish this dataset to enable other researchers to use the results for further studies.

---
*Corresponding author: Patrick Harms**, Nuremberg Institute of Technology, OHM-UX, Nuremberg, Germany,
E-mail: patrick.harms@th-nuernberg.de. https://orcid.org/0000-0002-9988-2028
Dominick Leppich, **Carina Bieber and Katrin Proschek**, Nuremberg Institute of Technology, OHM-UX, Nuremberg, Germany,
E-mail: dominick.leppich@th-nuernberg.de (D. Leppich),
bieberca73037@th-nuernberg.de (C. Bieber),
katrin.proschek@th-nuernberg.de (K. Proschek)
Ulf Schubert, DATEV eG, User Experience & Touchpoint Design, Nuremberg, Germany, E-mail: ulf.schubert@datev.de

In this paper, we present the process of this dataset creation and the design of the study where we gathered this data. In Section 2, we first explain basic terminology in the context of user interaction data and emotions and the tooling we used to capture these. Section 3 highlights the current state of research in this area and how our work integrates into existing work. Our approach to create the dataset of user interactions and emotions is described in Section 4. The resulting dataset is described in Section 5. Finally, we conclude this paper and give an outlook to further research based on this dataset in Section 6.

2 Foundations

In this section, we introduce basic concepts and terminology that is required to follow the process of the generation of our dataset.

2.1 User interactions

In event-driven UI technologies, the user interaction with the UI leads to the emission of *events* like key presses or mouse movements. Each event is of a certain type and contains a timestamp when this event was emitted. We group these event types into three categories: Keyboard event types, mouse event types and higher-level UI component event types (e.g. UI element focus). There are also event types for other input devices, but those are out of scope of this work.

We focus only on two keyboard event types: pressing and releasing of a certain key. Each key event contains the key which is pressed or released. For the mouse, we distinguish between six different types of events: Two for pressing and releasing the mouse button, two for single and double mouse clicks, one for mouse scrolling and one for mouse movements. The events of the first four types contain the information about the button that was pressed or released. The scrolling events contain the number of units in the direction the UI element was scrolled. Whenever a mouse is moved, there are multiple mouse movement events, one for each recognized intermediate positions of the cursor. All mouse events contain the position of the cursor in the application.

The last group of events contains higher-level UI events, that are emitted on certain interactions with UI elements. Whenever a UI element is focused, a focus event is emitted, as well as a focus loss event for the prior UI element if present. If the UI element holds content data, a content change event is emitted whenever the content changes. In case of dropdown menus and radio buttons, a change of the selected element also emits events of this type.

2.2 Emotions

Emotions can be defined as complex psychological experiences that involve a range of subjective feelings, physiological responses, and behavioral reactions [7]. There are many different types of emotions that have been identified, such as happiness, sadness, anger, fear, surprise, and disgust, among others [8]. However, the classification of emotions is often a matter of debate, as different scientists and scholars have different criteria for defining and categorizing emotions. Furthermore, emotions are not clearly regulated or governed by a single system in the human body, which makes the study of emotions a large and multifaceted field of research.

Nowadays, there are different strategies to detect the emotions of users. One way is called *facial coding* and can be done manually or automatically. In the automatic approach, an algorithm analyzes video data of the users' face to detect common patterns of facial expressions, which are known to express specific emotions. The knowledge about these patterns is stored in databases. Most facial coding software databases are based on actors who mimic emotions and not on people who express real emotions [[9], p. 2]. "[F]indings clearly demonstrate that trained actors and untrained participants express the same emotions differently" [[9], p. 12]. This is based on the fact that "[u]ntrained participants compared to trained actors showed substantially less intense facial expressions, which is in line with previous research" [[9], p. 10]. Therefore, "[Automatic Facial Coding (AFC)] may not yet be capable of detecting very subtle emotional facial expressions in contrast to other research methods" [[9], p. 12]. This is a problem because this means that the small expressions of emotions from people in everyday life cannot be recognized or are misinterpreted if, for example, the context is missing. Another problem is that joyful facial expressions are often better recognized by AFCs compared to unpleasant facial expressions [[9], p. 11]. The findings of Ulrich Föhl T. et al. [10] showed a lower sensitivity for unpleasant facial expressions. This finding is problematic, as unpleasant facial expressions are more important for the purpose of analyzing the UX of UIs.

Another way of measuring user emotions is the Semantic Differential (SD), a type of survey rating scale used for psychological measurement [11]. With the SD, Osgood has designed an instrument to capture the emotional mediation processes [[12], p. 130]. As our study was performed in German, we used an existing German SD from Stürmer et al. [13]. In our study, we let the users self-assess their emotions with the SD before and after the study, to verify the results of the detected emotions.

3 Related work

Measuring emotions reliably is a complex and challenging affair. Therefore, different studies on emotion recognition have been carried out. Overall, there are different approaches to detecting user emotions reliably. For example, some studies have used external body signals [14], internal physiological signals [15], and other contextual signals [16]. The recent research found evidence for the possibility of emotion recognition based on KMT dynamics data [[5], p. 2], smartwatch sensors [[17], p. 1], and facial expression, body posture, and gesture analysis [[18], p. 23–25].

Khanna and Sasikumar [19] performed an empirical study to measure keyboard stroke patterns. Thereby, Khanna and Sasikumar [19] used a self-assessment questionnaire for emotion recognition. They observed that participants followed the trend of decreased typing speed when experiencing negative emotions and showed a trend of increased typing speed when experiencing positive emotions. Compared to our work, Khanna and Sasikumar [19] considered only keyboard stroke patterns and did not induce emotions in advance. Our approach covers the induction of emotions and uses more than one data source.

Hibbeln et al. [20] performed three studies monitoring mouse cursor movements to show that mouse cursor movements are a real-time indicator of negative emotions. In the first two studies, they manipulated negative emotions and then observed the participants' mouse cursor movements. The third study was an observational study, without direct manipulation of the participants' emotions. Instead, participants had to report their level of emotion after each task. From their studies, they were able to conclude that negative emotions increase the distance and reduce the speed of mouse cursor movements. Furthermore, they found that the distance and speed of mouse cursor movements can be used to infer the presence of negative emotions and the level of negative emotions. In comparison to our work, Hibbeln et al. [20] considered only mouse cursor movements. In our approach, we also collect keystrokes.

Shikder et al. [21] performed a study collecting keystrokes from free text entry and mouse usage data after inducing different states of emotion through various multimedia components. In comparison to our study, Shikder et al. [21] had not verified the induced emotions in any way. In our study, we collect four different data sources in total. Two of these data sources are used to determine the users' emotions. The first data source is an external body signal, more precisely facial expressions. The second source consists of self-assessment questionnaires. The remaining two data sources are recordings of the usage of the keyboard and mouse interactions and do not contain information about the users' emotions. These four data sources form the basis of our dataset of user interactions and user emotions, which can be later used to look for dependencies between the interactions and emotions.

The base for our approach to recognizing emotions from keystrokes and mouse dynamics is the paper from Yang and Quin [5]. Yang and Quin [5] attest in their paper the possibility of emotion recognition from KMT dynamics by providing a summary of the scientific research in this field in the last decade. Overall, in the last few years, the research on unobtrusive emotion recognition increased, as the use of technologies like computers and smart devices increased both in the work environment and during leisure hours [[5], p. 3]. However, there are still gaps, which make further research necessary. One of the gaps is that most research focuses on one of the different possible measurement methods or even just one data source (e.g. [17, 20]). Another gap is that in most studies, there is no verified continuous emotion data available for the recorded KMT data as a reference, referred to as *ground truth*. Compared to our work, most studies focused on keystroke-related data only in combination with self-assessment questionnaires (e.g. [19, 22, 23]). In the area of mouse-related data and keystroke dynamics on touchscreen keyboards, in addition to self-assessment questionnaires, some studies also included induction of emotions (e.g. [20, 24]). Overall, in most studies, external body signals and internal physiological signals were not used for the ground truth, as the participants just filled in self-assessment questionnaires [[5], p. 13]. Especially ground truth data that detects the real emotions and their intensity, instead of just relying on the participants' information, is a crucial factor for the validity of the studies. The results of scientific research in this field have to be viable and accurate to achieve the goal of detecting the participants' emotions based solely on the users' interaction with a technical device. "Therefore, in the future, using biological signals as ground truth in parallel to the KMT dynamics should be a potential method that improves training data quality and help researchers evaluate their algorithm performance more objectively and accurately" [[5], p. 13]. In this project, rather than relying only on participants' self-assessments, we also work with external body signals to generate our ground truth following this recommendation. Furthermore, in our study, we attempt to induce different emotions first and then analyze KMT dynamics data to detect the emotions.

4 Approach

We created a dataset of user interactions tagged with information about the respective emotions. In this section, we describe our approach of the dataset creation process.

4.1 Experimental design

In order to create a dataset of interaction and emotion data, we had to perform a study with several user tests. We have defined recruitment criteria for our participants, set up a test application and tested the whole process in a pretest. The test application was instrumented to record user interactions on a low level of keystrokes and mouse movements. We also recorded additional sensor data to finally derive emotional data for each user test. We performed the study with 50 participants in our own laboratory. Afterward, we performed some post-processing steps for our dataset that were required after the study was completed. The participants were not told that we perform an emotional analysis beforehand, but instead that they will perform a user test of a test application. All these steps are described in more detail in the following sections.

4.2 Test application

The test application is a travel expense report that is implemented as a web application. It consists of seven pages with data entry fields of different types. There are one line text input fields, multi line text input fields, dropdown menus, checkboxes and groups of radio buttons. The first page queries the user for personal information like name and gender. On the second page, the user has to enter company internal data. The third page contains input fields for
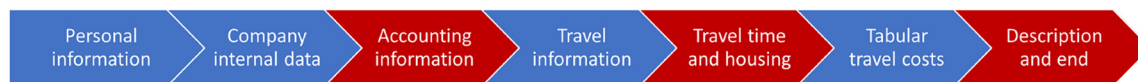
Figure 1: Structure of the travel expense report test application. The red pages contain emotional triggers.

Figure 2: First trigger on page three, which is a confirmation dialog indicating that the entered data is outdated but still correct to confuse the user.

Beginn

Falsches Format! Bitte Uhrzeit wie folgt eingeben: hh:mm:ss

Figure 3: Second trigger on page five, which is the input page for the temporal travel information. If the participant does not enter the required clock time format, an error message (red block) appears to request the required format.

accounting information. On the fourth page, travel information has to be entered. Page five queries the user to input time related travel information as well as transport and housing information. A detailed list of all travel costs can be entered in a tabular form on page six. The last page contains two larger text boxes to enter a description of the travel and additional notes, and two checkboxes to agree to the data processing and confirm the correctness of the entered data. Figure 1 shows the overall structure of the test application.

Three of these data entry pages contain emotional triggers that should cause different emotions of the users. The

triggers are located on the pages three, five and seven with at least one page without triggers in between. We chose to have pages without triggers in between to lower the emotional level to an average before and after the triggers.

The first emotion trigger is a confirmation dialog on the accounting information page. The dialog is shown in Figure 2 and indicates that the entered data is valid but outdated. With this trigger, we intended to confuse the participants slightly. The dialog can simply be closed to overcome this situation.

The second trigger is a clock input field that requires an unnecessarily complicated format by prompting for seconds and require zero padding, as shown in Figure 3. This trigger shall induce annoyance in the participants.

Finally, the third trigger simulates a system error and clears the contents of both text fields on the last input page. All input fields on the other pages are not affected by this simulated error. The last trigger is a strong negative trigger, shown in Figure 4 and shall induce strong negative emotions.

The triggers are set to contradict the standardized interaction principles from ISO 9241 Part 110 [25]. For example,
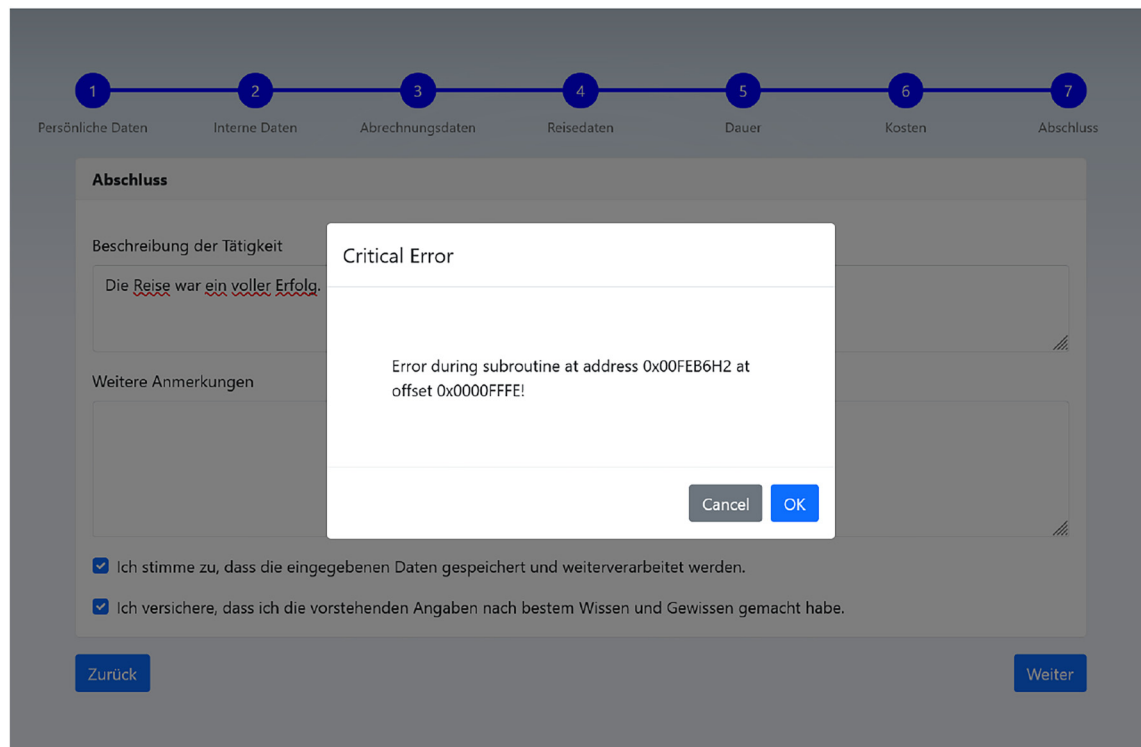

Figure 4: Third trigger on page seven, which is a simulated system error that clears the contents of both text fields on the last page. When the participant clicks next, the contents of the page are darkened and a cryptic error message appears. After closing this error message the contents of the text fields are deleted.

task appropriateness, self-descriptiveness, conformance to expectations, and avoidance of user errors are lacking when the system prompts for an unnecessarily detailed time format. Furthermore, the required time format is not presented beforehand. The triggers mainly target negative emotions, as negative emotions are a sign of bad UX and our final goal is an automated UX evaluation and improvement. A detailed overview of all seven pages of the travel expense report can be found in the Appendix A.

The test application is embedded into a web application that we have developed for this study. The other parts of the web application are a greeting page, an introduction page, two pages for a user self-assessment, two feedback pages and a final end page. These other pages are shown in the Appendix B.

4.3 Study preparation

As part of the preparations for the study, we defined recruitment criteria for study participants for the selection of the target group. These were to fit both the sample application of a travel expense report and the actual goal of creating the dataset. Table 1 shows the criteria that we derived for this purpose. It is important to note that we explicitly excluded test subjects with IT, design, or psychological backgrounds. The reason is that these careers could influence the study results as e.g. psychologists may realize that emotions are being recorded. On this basis, we recruited the subjects.

Furthermore, we prepared all documents required for the study. For data protection reasons, it was necessary to have the test persons sign a data protection declaration. We also prepared a receipt template for the handover of an expense allowance, which was also signed by the subjects. In order to standardize the study process between all moderators, we created a written guideline that describes all steps of the study, including technical set up and execution. In addition, we defined a uniform welcome text, that we read out at the beginning of the session. By defining the text beforehand, we wanted the test sessions to be as similar as possible.

We defined the input information which the participant had to insert to avoid capturing personal data of the participants. The participant got access to the input data through printouts which were fixed to the laptop where we conducted the study. Sometimes, the participant had to assign information or decide which one to insert.

4.4 Tooling

To record the user test sessions and capture all user interaction and emotion data, we have used two existing software tools. We used the Automated Quality Engineering of Event-driven Software (AutoQUEST) tool suite to record the users' interaction [26]. AutoQUEST contains monitor components to record interaction data in different event-driven UI technologies (e. g. Java, Android, Web, Mixed Reality (XR)). We have already used AutoQUEST multiple times in earlier works to gather interaction data [27, 28, 29]. There are also extensions to generate usage profiles and perform automated usability evaluations based on the recorded interaction data [27].

The software iMotions was used to record the user test session together with data from multiple connected sensors and devices. IMotions can "[a]ssess emotions in facial expressions by [u]sing automated facial coding" [30] and is "[a] single software for all your human behavior research" [31]. It is possible to connect different devices with iMotions and add various sensors that can be captured. For example, a Tobii Pro Fusion eye tracker or a webcam can be used for a recording of the test subjects faces. Every recording contains a video capture of the webcam, a screen capture, eye tracking data and basic keyboard and mouse interaction data. IMotions contains an AFC module called "Affectiva's AFFDEX and Realeyes" that is capable of analyzing the test subject's facial expressions to predict emotions. The supported emotions are: Anger, Confusion, Contempt, Disgust, Engagement, Fear, Joy, Sadness, Sentimentality, Surprise and Valence. We used this module to automatically predict intensities for all supported emotions at any given time during the test session. The assignment is independent of the external appearance or the presence of accessories like glasses.

Table 1: Test subject recruitment criteria.

Criteria	Stipulating
General description	Persons who are or were working with experience in travel expense accounting or with the possibility that they will deal with travel expense accounting in the future.
Age	18–65 years old.
Career	Every job that involves administration tasks on the computer. The following jobs are excluded: All kinds of IT jobs, psychologists, psychiatrists, designers.
IT experience	Regular work with a computer.

4.5 Recording of interaction data

In order to capture the user interaction in our study, we used a Hyper Text Markup Language (HTML) monitor of the AutoQUEST tool suite. The HTML monitor is a standalone server application. It provides a JavaScript that can be added to websites to automatically connect to the monitor application and to send all preconfigured events of all relevant HTML elements of the websites to the monitor. We integrated this JavaScript into our web application. We set it up to capture events on individual HTML elements as well as globally on the website. Furthermore, we capture single mouse clicks on all individual HTML input elements like text fields, multi-line text areas, checkboxes, radio buttons and dropdown menus. When appropriate, we also captured events if the content of these HTML elements change or if the HTML element gets or loses the focus. Finally, we set it up to capture key press and release events as well as mouse clicks, double clicks, mouse button press and releases, scrolling and mouse movement events globally on the website.

After setting up the script and embedding it into our web application, we captured the user interaction automatically when the user interacted with the website. We started recording the interaction already on the first welcome page and stopped when the user closed the web browser after visiting the final end page.

For each user test session, the HTML monitor created a log file with all interactions together with their timestamps. When exiting data entry, we additionally saved the values of all final field entries along with a timestamp. In case of the simulated system error (see Figure 4), we saved before and after the form reset.

It is still very common in working environments to work with the keyboard and mouse and this type of interaction was also most relevant to our cooperation partner. Therefore, the study was performed on a laptop with an external Bluetooth mouse. The touchpad and touch screen were disabled to enforce the usage of the external mouse. We reduced the number of input devices to guarantee a larger dataset of keyboard and mouse interaction data. We plan to include other types of input devices like touch screens in future studies.

4.6 Recording of emotion data

We used the software iMotions to organize and perform the study and to capture additional sensor data during the user sessions. We used a laptop with an integrated webcam as hardware for the study. Overall, we collected and combined different types of data, as we needed them to detect both the emotion and the source of the emotion in the user interface. During the user sessions, the screen content was recorded as a video file, the users were recorded with the webcam, as well as basic mouse and keyboard interactions were captured. The emotions of the participants were captured with multiple approaches. One approach is the emotion detection based on facial coding. The webcam recording was used by iMotions' Affectiva AFFDEX module to detect the users' emotions automatically. Additionally, we asked the participants to self-assess their emotions before and after the test application without explicitly telling them about our reasoning.

At the beginning of the study, we needed to level the participants' emotions before they started using the test application. We used a film clip to level the participants' emotions. We chose the opening scene from the movie 'Pride and Prejudice' from 2005. The scene has a duration of 1 min and 31 s. The film clip induces calmness. We use this film clip based on the article "Ratings for emotion film clips" by Gabert-Quillen et al. [32]. In a previous work, we elaborated on this stimulus material [33].

After the participants' emotions have been brought to a consistent level, we asked them to fill a sentiment query using the SD. This SD consisted of six bipolar pairs of adjectives, each rated on a nine-point scale, as shown in Figure 5. The results of the SD is one of our approaches to assess the participants' emotion.

In our test application, we have integrated the three emotional triggers mentioned in Section 4.2. To ensure these triggers have the intended effect, we have created two versions of the online form, one with and one without the triggers. We split the participants into two groups, one group using the test application with the triggers and the other group without the triggers.

After the users have finished using the test application, a 5-star feedback and an expectation survey were carried out, followed by a second SD. Finally, post-session questions were asked in the form of a short interview. The questions of this interview were prepared beforehand and had a special focus on the triggers and the participants' emotions while encountering them. Depending on whether the session was with or without triggers, some of the questions were omitted. In the process, questions were asked about what the respondent noticed positively or negatively, and the triggers and the emotions they triggered were discussed in more detail. The results of the interview were recorded in writing and filed and were later used to verify the results of the facial coding emotion detection. The original German interview questions for the group with triggers are shown in Table 2 in the Appendix C.

Wie ist Ihre aktuelle Stimmung?

unzufrieden	○ ○ ○ ○ ○ ○ ○ ○ ○	zufrieden
unglücklich	○ ○ ○ ○ ○ ○ ○ ○ ○	glücklich
genervt	○ ○ ○ ○ ○ ○ ○ ○ ○	erfreut
verzweifelt	○ ○ ○ ○ ○ ○ ○ ○ ○	hoffnungsvoll
schwermütig	○ ○ ○ ○ ○ ○ ○ ○ ○	ausgeglichen
gelangweilt	○ ○ ○ ○ ○ ○ ○ ○ ○	entspannt

Weiter

Figure 5: The SD to assess the participants emotions before and after the usage of the test application.

4.7 Execution of the study

The study was tested in a pretest beforehand and changed accordingly after reviewing the pretest results. Due to these adjustments, we have not experienced major issues during the study execution.

The study was conducted from October 6 to November 4, 2022, at the OHM User Experience Center (OHM-UX) of the Nuremberg University of Technology. We selected a total of 50 test subjects for the study using the established recruitment criteria in Table 1, of which 22 were male and 28 were female. The group consisted of professionals and students who had experience with travel expense reporting or may have to report travel expenses in the future. The age range was from 21 to 64 years old. None of the subjects had heard of the project or used the application before.

For the study, we divided the subjects into two groups. Group V0 included 11 subjects and conducted the study without triggers. Group V1 included 39 subjects and received the form with the three negative triggers. This split allowed us to assess the average emotional level during the interaction with the travel expense report and to check if there were any additional problems with the online form besides the intentionally set negative triggers. The time window for the study was 1 h per subject. The subjects were not told that an attempt would be made to record their emotions, as most people behave differently when they know they are being watched [34] and since we did not want to influence the results. Instead, it was communicated that this was a test of the usability of the form to achieve the goal of having subjects show their emotions naturally and not falsify them.

In the following, the procedure of a user session is described in detail. We prepared the room and the hardware as defined in our guideline (see Section 4.3). After the subject arrived and was greeted, we read out the prepared welcome text and answered all questions of the participant. After signing the privacy statement, we started the study in iMotions and left the test person alone in the room. A one-way mirror made it possible to observe the test persons without distracting them during the test. First, the film clip was shown to the test person to level their emotional level (see Section 4.6). Afterward, the test person was asked to fill the first SD. After completing the survey, the introductory text was shown to the user that explains the setting of the fictional test situation. Then the travel expense report application opened and the test subject started using it. After finishing data entry into the test application and submitting the form, the second SD was conducted. This was followed by a 5-star feedback and an expectation survey. When the user submitted the feedback, the technical part of the test session was over, and we went back into the room and performed the post-session interview. We explained, that the study was not about a user test of the test application but about the analysis of emotions. Finally, the handing over of an allowance, the signing of the receipt and the farewell of the test person took place. The participants did receive 50 Euro compensation for participating in the study.

4.8 Conclusion of test sessions

The evaluation showed that five of the 50 datasets had problems. One of them was an unrecoverable error in iMotions

that corrupted the dataset. In one case there was an unintentional influence of a test person, therefore we decided to remove this dataset entirely. Two datasets were accidentally recorded into one log file, and manual work was required to split the data into two recordings. A total of 46 datasets could be used without further restrictions. These included 36 sessions with triggers (V1) and 10 sessions without triggers (V0).

4.9 Emotion data tagging

The analysis of the facial coding exposed three main challenges. The first one is related to facial expressions. Mainly as "[facial expression recognition] provides valid measures for most emotion categories if […] participants are […] in a typical lab setting with frontal face recording and good lighting condition" [[35], p. 2]. Although we tried our best to achieve this, we still faced diverse challenges (e.g. the participants did not continuously look straight into the camera). We already mentioned previously that most facial coding algorithms were trained based on mimics of actors. This leads to issues when detecting subtle mimics and the fact, that positive emotions are detected easier in general. An example could be a smile as part of an expression of disbelief or confusion resulting from a negative experience, such as the last trigger of the simulated system error. As the software does not know the context and seemingly also does not take into consideration the participants' emotions in the last few seconds, this facial expression sometimes evaluates as joy and confusion. The last issue is a result of the post-session questions, where the participants told us about their feelings during the interaction with the test application. Therefore, we know what they have experienced during a certain time of interaction with the form. As we compared the provided information with the facial coding of the participants, sometimes the information and the result of the facial coding did not match. For example, peaks are missing, or the facial coding provided us with a different emotion.

Due to these mentioned issues, we could not solely rely only on facial expression recognition and could not blindly trust the classification of the algorithm. Consequently, we decided to annotate the emotions manually.

4.10 Manual annotation of emotions

For the manual annotation, three individuals independently analyzed the dataset, each using their own methodology. The methods were not discussed among each other in detail to avoid mutual interference. Table 4 in the Appendix D describes the three methods in detail.

In order for the results to be in the same format and thus easier to subsequently compare and combine, we defined certain guidelines. These included the common definition of labels for emotions. The emotions considered were Anger, Confusion, Contempt, Disgust, Engagement, Fear, Joy, Neutral, Sadness, Sentimentality, Surprise and Judgement. The naming of the emotions was analogous to the naming of the emotions in the iMotions software. Only in the case of the emotion "Valence" was a change of meaning made to the emotion "Judgement", since this naming better matched our assessment of this emotion. In addition to the naming of the emotions, a percentage for the strength of the emotion was added. Here we limited ourselves to the values 33, 66 and 100, since it is difficult to manually distinguish between the emotions more precisely.

All sessions were first divided equally among the three evaluators, and the emotions were annotated in the iMotions software according to the respective methodology. Subsequently, a rotation of the sessions to be annotated took place and each session was evaluated by a second person. Finally, there was a final rotation and the third person reviewed the evaluation of the previous two evaluations. Due to the use of different methods, there were variations in the annotations, for example in emotion density. However, the division and rotation ensured that the final result was not shaped by one particular method, but was an overall result from all three annotation methods. To ensure this also in the third step, a set of rules was defined to ensure a uniform procedure. It was defined how to proceed if an emotion was only annotated in one run, the time span deviates, another or a new emotion is detected. Simply speaking, only the emotions with their values on which the majority of the three evaluators agreed were finally annotated. The complete rule set can be seen in the Appendix D in Table 5.

Finally, the results of this manual annotation were exported from iMotions in the form of a CSV list containing, among other things, a timestamp and the annotated emotions.

4.11 Data post-processing

For further analysis, we anonymized the dataset as far as possible. All personal data or names were replaced by a subject number from 1 to 50. An exception were the videos, which were needed for further analysis and therefore could not be anonymized. These videos are not part of the published dataset. In some test sessions, the participants entered personal data instead of the given fictional data into some of the input fields. These mistakes were always recognized but the respective keyboard events containing the personal information were still captured in

these cases, we manually removed the information about which key was pressed from the dataset.

Overall, we have two main data sources, KMT data from AutoQUEST and emotion data from iMotions including the emotions from the AFC as well as the manual annotations. In addition, we have as data sources the final field assignments that our web application saved to a database. Most importantly, we ensured that all three data sources collected the same data regarding the content of the interaction elements in the online form.

To create a dataset containing both, the interaction data and emotion data, we had to combine the interaction and emotional data from the two main data streams. The main challenge was the different format of timestamps in AutoQUEST and iMotions. In order to assign each interaction the correct emotions, we had to align both data sources with respect to the time. The reason for the alignment is that AutoQUEST stores its timestamps as Unix time stamps with an accuracy in the millisecond range, whereas iMotions uses a different encoding for the time stamps. We have not found any documentation that describes the data in more detail. Nevertheless, the time stamps are not Unix time stamps. It seems to be a floating point number describing the milliseconds since the start of the recording. We also noticed that the start time of iMotions is different from the first event recorded with AutoQUEST. Therefore, we needed a strategy to create a correct temporal alignment. Both the iMotions and AutoQUEST recordings contain keyboard interaction data, which allowed keystroke events appearing in both files to be defined as times of temporal matching. Subsequently, the two datasets were overlaid in such a way that the deviation between each matching keystroke event was minimal. The result of merging the data is a dataset of user interaction data for which values for each emotion are simultaneously stored for each interaction. This dataset includes both the automatically detected emotions by iMotions' facial coding algorithm and our manual annotations.

5 Dataset

The result of the study is a dataset containing keyboard and mouse interaction data together with detected emotional data from iMotions facial coding module and manually annotated emotions. We chose a human readable CSV file format for the dataset to be easily usable with state-of-the-art ML software. We tried to record all possible types of interactions for the keyboard and mouse in a web application to make this dataset as unconstrained as possible. Our dataset consists of two files: `v0.csv` and `v1.csv`. The file `v0.csv` contains the interaction and emotion data for all 10 test sessions with the triggers being disabled. The file `v1.csv` contains the same data for all 36 test session with the triggers being enabled. Both files are structured the same way. Each row of the file contains tab separated values. The first row of the file is the header and describes what type of data is present in the respective column in the following rows. Beginning from the second row, each row represents a recorded user interaction with all its parameters and emotion annotations both from the iMotions Affectiva module and our manual annotation. The following listing describes the meaning of every column in the dataset.

The dataset can be downloaded at: https://doi.org/10.5281/zenodo.7778612.

session This value represents the test session the event belong to. All sessions are numbered from 1 to the last one of the respective group, and all events in a session are ordered chronologically.

timestamp This value corresponds to the Unix timestamp of the event emission. The resolution of the timestamp is in milliseconds.

type This value indicates what type of event it is. The possible types are: `KeyPressed`, `KeyReleased`, `MouseButtonDown`, `MouseButtonUp`, `MouseClick`, `MouseDoubleClick`, `Scroll`, `MouseMovement`, `TextInput`, `ValueSelection`.

target This value indicates from which UI element the event was emitted. The value contains a hierarchy of UI elements, with the element that emitted the event being mentioned first. Some UI elements can have an `id` that was uniquely defined for our test application. All relevant UI elements of the travel expense form have an associated ID.

alt This value is only present for keyboard related event types and is `TRUE`, if the alt modifier key is being pressed in parallel to the respective key event and `FALSE` otherwise.

control This value is only present for keyboard related event types and is `TRUE`, if the control modifier key is being pressed in parallel to the respective key event and `FALSE` otherwise.

shift This value is only present for keyboard related event types and is `TRUE`, if the shift modifier key is being pressed in parallel to the respective key event and `FALSE` otherwise.

meta This value is only present for keyboard related event types and is `TRUE`, if the meta modifier key is being

pressed in parallel to the respective key event and FALSE otherwise. The meta modifier key is the Windows key on most keyboards.

key This value is only present for KeyPressed and KeyReleased event types and describe the key being pressed or released. Some participants accidentally typed in their personal information at some point during the test session. In these cases, we have replaced the key with the constant ANONYMIZED.

repeat This value is only present for keyboard related event types. If the event is emitted multiple times due to the key being hold pressed, the first emission has the repeat value FALSE and all following events the value TRUE. In all other cases, the value is FALSE.

button This value is only present for mouse button related event types and indicates which mouse button is being pressed or released. Possible values are: LEFT, RIGHT and MIDDLE. However, the value MIDDLE is not present in our dataset.

x This value is only present for mouse related event types. It corresponds to the *x* coordinate of the mouse cursor in the application when the event was fired.

y This value is only present for mouse related event types. It corresponds to the *y* coordinate of the mouse cursor in the application when the event was fired.

xPosition This value is only present for the Scroll event type. It corresponds to the amount of units scrolled horizontally.

yPosition This value is only present for the Scroll event type. It corresponds to the amount of units scrolled vertically.

selectedValue This value is only present for ValueSelection event types which are emitted, if the value of a dropdown menu or a radio button group is changed. The value represents the new value being selected.

enteredText This value is only present for TextInput event types, which are emitted if the content of UI elements is changed. This value represents the new content value of the UI element.

Emotional data is present for 12 different emotions. The automatically detected emotions by iMotions facial coding module are: Anger, Confusion, Contempt, Disgust, Engagement, Fear, Joy, Neutral, Sadness, Sentimentality, Surprise and Valence. The manually annotated emotions are the same, except for the Valence emotion that was replaced by Judgement. There is a column for every emotion for both sources, the facial coding and manual annotation, resulting in a total of 24 columns.

emotion_affectiva_EMOTION These 12 columns contain the detected emotion intensities for the emotion EMOTION, detected by iMotion's Affectiva module. The values range from 0 percent to 100 percent.

emotion_manual_EMOTION These 12 columns contain the manually annotated emotions. The values range from 0 percent to 100 percent.

6 Summary

The goal of this work was to create a dataset of interaction data with annotated emotional data that can be used for ML algorithms. We have performed a study with 50 participants and a test application of our own. The results of 46 of the 50 test sessions were usable without any issues and were transformed into our final dataset. We have recorded all possible types of interactions and exported the dataset in a format that can be used with state-of-the-art ML software. With this dataset, it is possible to further research relationships between user interactions and user emotions. We hope that other researches can use our results for their own studies in this area.

Our next plan is to evaluate the results and look for possible relationships ourselves. If we manage to find simple rules of interaction behavior that leads to certain emotions, we could create a rule based algorithm to predict the emotions. Otherwise, we can calculate metrics on this interaction data and use the results as the input for machine learning algorithms to train a predictor for the emotions. These prediction algorithms need to be evaluated with respect to their accuracy.

Appendix A: Travel expense report test application overview

Figures 6–12

Figure 6: First page of the travel expense report test application. It queries the title, first and last name, and the gender from the user.

Figure 7: Second page of the travel expense report test application. It queries contact information as well as company internal data from the user. There are multiple types of contact information that can be entered: The email address, the mobile phone number, the address and a field for other information. Every contact information has a checkbox which displays a text input field below if activated. The company internal data are: the area of the user, the position, the staff number, the superior, and the area director.

Figure 8: Third page of the travel expense report test application. It queries accounting information from the user. The accounting information are: the accounting contact person, the area, the area number, the accounting code, and the travel id.

Figure 9: Fourth page of the travel expense report test application. It queries information about the travel from the user. This information contain the type of travel (business travel or seminar/congress/further training), a description of the goal of the travel, the travel destination, and the organizer.
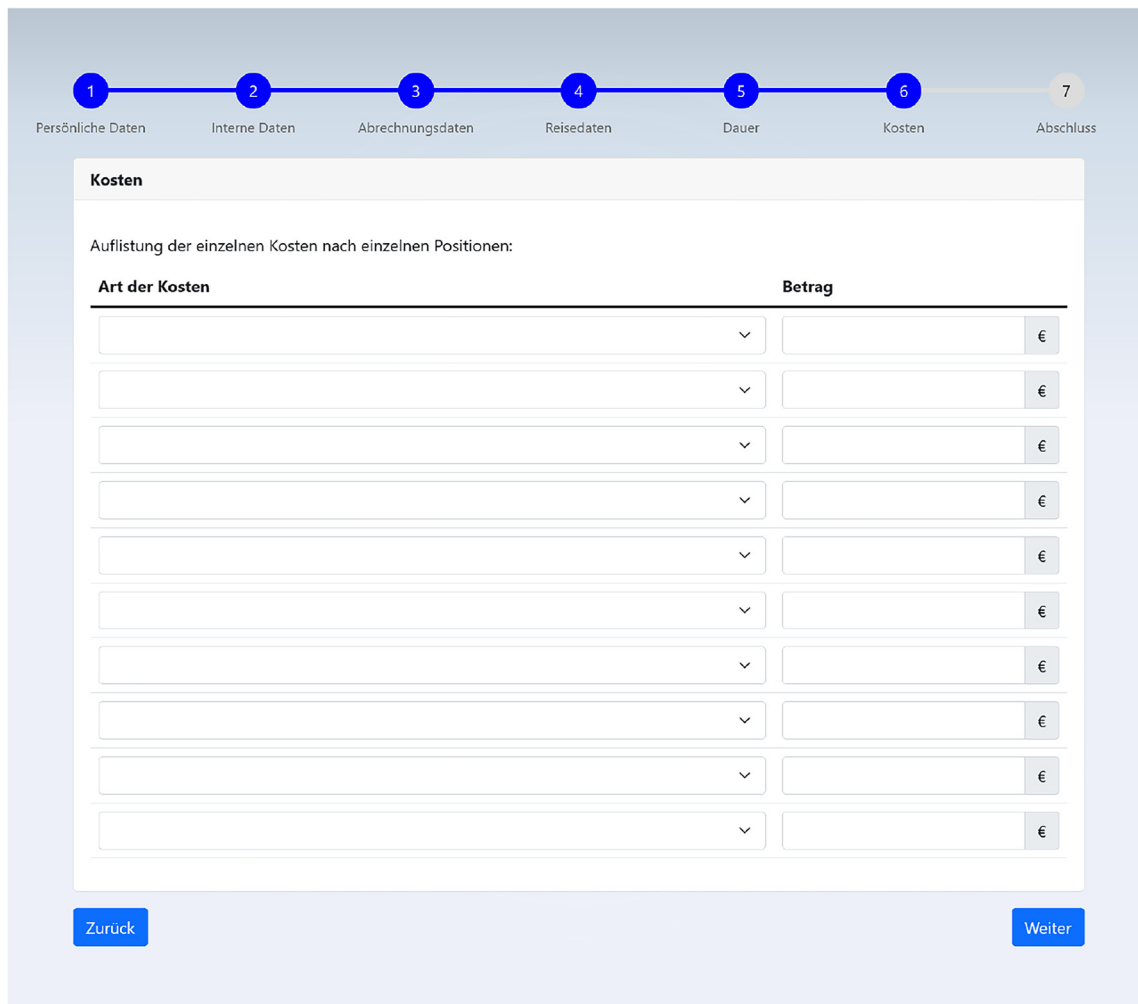
Figure 10: Fifth page of the travel expense report test application. It queries the duration and travel related information from the user. The following information has to be entered: the duration of the travel (in hours, days, or weeks), the beginning and end of the travel (date and time), the types of transport used in the travel (public transport, plane, and car), and the housing.

Figure 11: Sixth page of the travel expense report test application. It queries a detailed table of all travel related costs. The page contains a table with ten rows and two columns. The first column refers to the type of costs and contains a dropdown menu with a predefined set of cost types. The second column refers to the amount of money spent on the corresponding exposure and contains an input field to enter this amount.

Figure 12: Seventh page of the travel expense report test application. It queries two additional free-text information about the travel and contains two checkboxes and the final submit button. The first text input field queries a description of the activity of the travel. The second text field can be used to enter any kind of additional remarks. The first checkbox has to be checked to agree to the data processing. The second checkbox has to be checked to confirm that all entered data is correct. When both checkboxes are checked, the final submit button gets enabled and the travel expense form can be submitted.

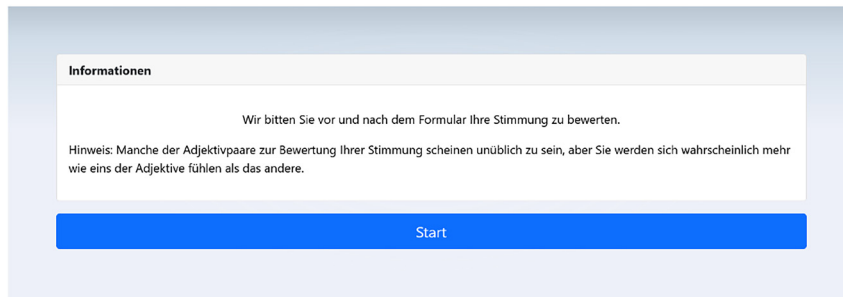# Appendix B: Additional pages of our web application

Figures 13–17

Figure 13: Initial page of the web application. It contains a "Start" button to start with the data entry and asks the user to self-assess their mood on the following page and after the interaction with the test application.
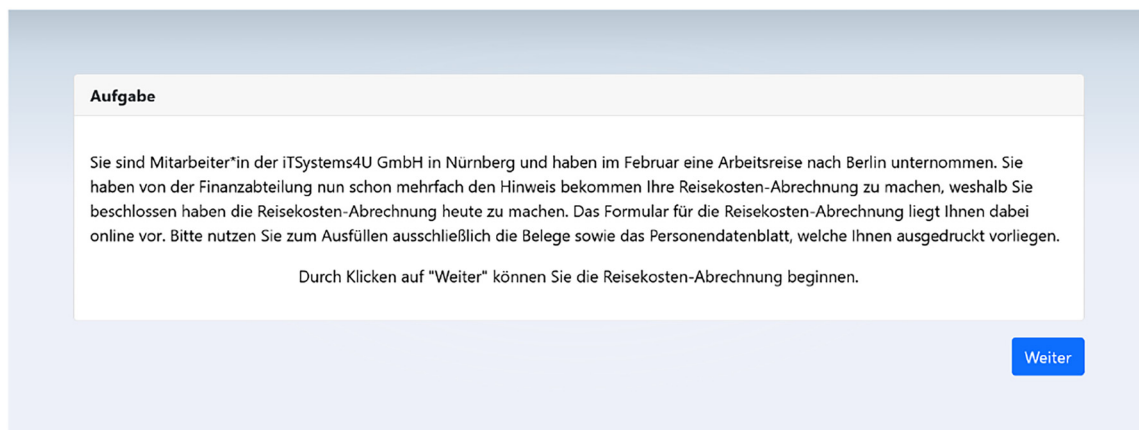

Figure 14: Third page of the web application. It directly follows the first SD and introduces the participant to the test scenario. The user acts as an employee of a fictional company and has done travel to Berlin back in February. The financial department has asked to perform the travel expense report, therefore the users does it now. The form of the travel expense report is available online. The user is asked to exclusively use the data which is printed out.

Figure 15: Page eleven of the web application. It queries the user for feedback on the test application. The user shall rate their satisfaction on a 5-point scale from very dissatisfied to very satisfied. Below there is a text input field which shall be used to justify the rating.

Figure 16: Page twelve of the web application. It queries the user for additional feedback on the test application. The users shall rate five different topics with respect to the expectation being not satisfied, satisfied or exceeded. For each topic, there is an additional option indicating that this topic was not expected or was not relevant to the user. There is a sixth topic row that can be freely specified and rated by the user.
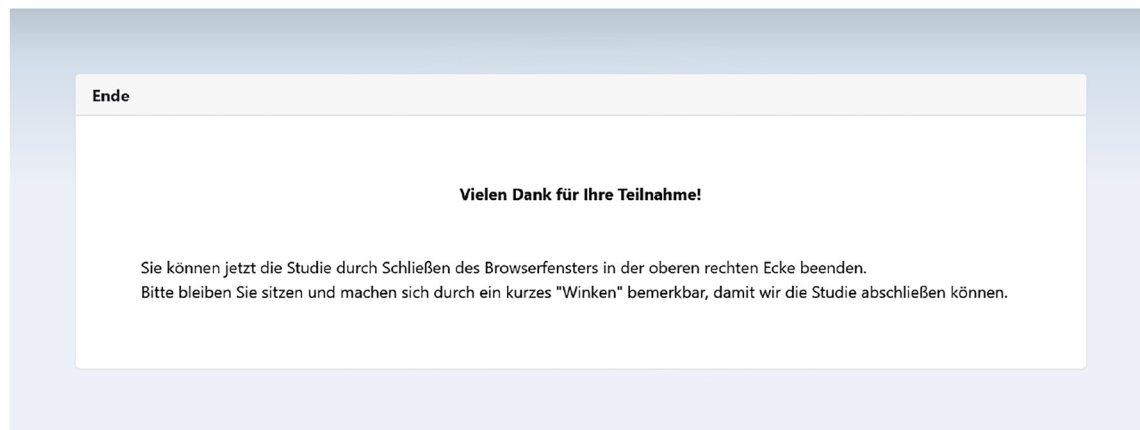
Figure 17: Page fourteen of the web application. It directly follows the second SD. It thanks the user for participating in the study and asks the user to close the web browser. Afterward, the user shall give a signal to the moderator.

Appendix C: Post-session interview

Table 3

Table 2: Original German post-session interview questions. The questions 3–5 were only asked for participants who were in the group with enabled triggers.

Nr.	Frage
1	Welche zwei bis drei Dinge mochten Sie besonders an diesem Formular?
2	Welche zwei bis drei Dinge müssten am ehesten verbessert werden?
3	Wie haben Sie sich gefühlt als bei den Abrechnungsdaten die Bestätigungsnachricht kam?
4	Wie stehen Sie zu der hohen geforderten Genauigkeit bei der Uhrzeit?
5	Wie erging es Ihnen als das technische Problem auftrat?
6	Wollen Sie uns noch etwas mitteilen?

Table 3: Translated post-session interview questions. The questions 3–5 were only asked for participants who were in the group with enabled triggers.

No.	Question
1	What two to three things did you like most about this form?
2	What two to three things would most likely need improvement?
3	How did you feel when the confirmation message came with the billing data?
4	How do you feel about the high accuracy required for the time?
5	How did you feel when the technical problem occurred?
6	Is there anything else you would like to share with us?

Appendix D: Manual emotion annotation strategy

Table 4: Methods for manual emotion annotation.

No.	Method
1	First, the observation was done without recording the person. Instead, the emotions automatically assigned by iMotions and the resulting graphs were considered, and their values were compared with the user's interaction with the form (mouse movements, speed, errors, triggers). Markers were set according to each. Subsequently, the graph excursions were viewed together with the camera recording of the test persons. The emotional state of the test persons was interpreted. The recording of the form was hidden, and it was compared which emotions were interpreted by iMotions and how the emotions of the test person were perceived by the evaluator. Then a marker was set with the corresponding emotion. Finally, all elements were faded in, and the recording was analyzed chronologically. This means that all events were considered, both the form, the subject, the camera recording and the graph. Here all sources were considered and the respective emotion of the test persons were interpreted and marked. Finally, the questionnaire was looked at to see if there was important information that could help with emotion recognition. If necessary, the respective passage was looked at again and markers were set anew.
2	In this method, more and more data streams were gradually added to validate the emotions. At first, only the recordings of the subjects' faces were analyzed at the discretion of the evaluator, then the audio recording was analyzed separately if necessary, then the emotions from iMotions were faded in to validate previous results and examine specific peak locations, and then the emotion annotation was aligned with the protocol and the free text field.
3	The focus of this method was to study facial expressions including micro expressions as accurately and objectively as possible with simultaneous observation of user behavior through eye tracking data, mouse movement, and text input. The evaluation of facial expressions relied predominantly on the Facial Action Coding System (FACS), a technique for measuring facial behavior developed in 1978 by Ekman and Friesen [8], and on personal perception. At the end, emotion triggers and the parts that were mentioned as problematic in the protocol were looked at more closely and checked again for emotions.

Table 5: Majority voting rule set for manual emotion annotation.

Situation	Rule
One emotion was only annotated by A or B	C decides whether this emotion is perceived. If yes, it remains. If not, its intensity is weakened or deleted. In case of doubt or uncertainty, the emotion should rather remain than be deleted.
Different emotions were annotated by A and B at the same time	Both emotions remain, or C decides if from his own point of view only one of the two emotions exists.
Different intensities of the same emotion were annotated by A and B at the same time	C decides according to its own discretion and methodology.
The interval of the emotion of A and B is different, but the emotion is the same	C decides at its own discretion on the start and end of the interval. In case of doubt, the longer interval has priority.
C recognizes an emotion that was not recognized by either A or B	The emotion is not annotated because there is no majority.
C does not recognize an emotion that was recognized by A or B	The emotion is weakened or deleted because there is no majority.
C does not recognize an emotion that was recognized by A and B	The emotion persists because the majority believes that the emotion exists.
A or B have annotated one emotion, but C sees another emotion	The emotion persists because the majority believes that the emotion exists.

References

1. Barnard A. *Die menschliche Seite der Digitalisierung*, 2022. https://www.siemens.com/de/de/unternehmen/stories/forschung-technologien/folder-topics/user-experience-for-digital-transformation.html (accessed 08, 2022).

2. Hamm. *Digitale User-Experience wird zum wichtigsten Unterscheidungsmerkmal im Marketing*, 2022. https://www.digital-verbunden.net/aktuelles/news/news-detail/digitale-user-experience-wird-zum-wichtigsten-unterscheidungsmerkmal-im-marketing/ (accessed 08, 2022).

3. UXQB. *CPUX-F Curriculum and Glossary*, 2020. https://uxqb.org/en/documents/cpux-f-en-curriculum-and-glossary-3-16/ (accessed 03, 2023).

4. Mishra K. *Challenges Faced by Facial Recognition System*, 2020. https://www.pathpartnertech.com/challenges-faced-by-facial-recognition-system/ (accessed 08, 2022).

5. Yang L., Qin S.-F. A review of emotion recognition methods from keystroke, mouse, and touchscreen dynamics. *IEEE Access* 2021, *9*, 162197−162213.

6. Ekman P., Friesen W. V. *Facial Coding Action System (FACS): A Technique for the Measurement of Facial Actions*; Consulting Psychologists Press: Palo Alto, CA, 1978.

7. Hockenbury D. H., Hockenbury S. E. *Discovering Psychology*; Macmillan, 2010.

8. Ekman P., Friesen W. V. Facial action coding system. *Environ. Psychol. Nonverbal Behav.* 1978.

9. Höfling T. T. A., Alpers G. W., Büdenbender B., Föhl U., Gerdes A. B. M. What's in a face: automatic facial coding of untrained study participants compared to standardized inventories. *PLoS One* 2022, *17*, e0263863. https://doi.org/10.1371/journal.pone.0263863.

10. Föhl U., Höfling T. T. A., Gerdes A. B. M., Alpers G. W. *Read My Face: Automatic Facial Coding versus Psychophysiological Indicators of Emotional Valence and Arousal*, 2020.

11. Zakharenko A. *Semantic Differential Scale: Definition, Questions, Examples*, 2020. https://aidaform.com/blog/semantic-differential-scale\ignorespaces-definition-examples.html (accessed 03, 2022).

12. Woll E. *Empirische Analyse emotionaler Kommunikationsinhalte von Printwerbung*; Deutscher Universitätsverlag: Wiesbaden, 1997; pp. 171−221.

13. Stürmer R., Schmidt J. *Erfolgreiches Marketing durch Emotionsforschung: Messung, Analyse, Best Practice*, Vol. 395; Haufe-Lexware, 2014.

14. Castellano G., Kessous L., Caridakis G. Emotion recognition through multiple modalities: face, body gesture, speech. In *Affect and emotion in human-computer interaction*; Springer, 2008; pp. 92−103.

15. Shu L., Xie J., Yang M., Li Z., Li Z., Liao D., Xu X., Yang X. A review of emotion recognition using physiological signals. *Sensors* 2018, *18*, 2074.

16. Ortega M. G. S., Rodríguez L.-F., Gutierrez-Garcia J. O. Towards emotion recognition from contextual information using machine learning. *J. Ambient Intell. Hum. Comput.* 2020, *11*, 3187−3207.

17. Quiroz J. C., Geangu E., Min H. Y. Emotion recognition using smart watch sensor data: Mixed-design study. *JMIR Mental Health* 2018, *5*, e10153.

18. Dzedzickis A., Kaklauskas A., Bucinskas V. Human emotion recognition: review of sensors and methods. *Sensors* 2020, *20*, 592.

19. Khanna P., Sasikumar M. Recognising emotions from keyboard stroke pattern. *Int. J. Comput. Appl.* 2010, *11*, 1−5.

20. Hibbeln M. T., Jenkins J. L., Schneider C., Valacich J., Weinmann M. How is your user feeling? Inferring emotion through human-computer interaction devices. *Mis Quarterly* 2017, *41*, 1−21.

21. Shikder R., Rahaman S., Afroze F., Islam A. B. M. A. A Keystroke/mouse usage based emotion detection and user identification. In *2017 International Conference on Networking, Systems and Security (NSysS)*; IEEE, 2017; pp. 96−104.

22. Charles Epp C. *Identifying emotional states through keystroke dynamics*. Ph.D. Dissertation, Citeseer, 2010.

23. Nahin A. F. M. N. H., Alam J. M., Mahmud H., Hasan K. Identifying emotion by keystroke dynamics and text pattern analysis. *Behav. Inf. Technol.* 2014, *33*, 987−996.

24. Trojahn M., Arndt F., Weinmann M., Ortmeier F. Emotion recognition through keystroke dynamics on touchscreen keyboards. In *ICEIS*, 2013; pp. 31−37.

25. DIN EN ISO 9241-110:2020. *Ergonomics of Human-System Interaction - Part 110: Interaction Principles*, 2020.

26. Herbold S., Harms P. AutoQUEST—automated quality engineering of event-driven software. In *2013 IEEE Sixth International Conference on Software Testing, Verification and Validation Workshops*; IEEE, 2013; pp. 134—139.

27. Harms P. *Automated Field Usability Evaluation Using Generated Task Trees*, 2016.

28. Harms P. Automated usability evaluation of virtual reality applications. *ACM Trans. Comput. Hum. Interact.* 2019, *26*, 1—36.

29. Harms P., Grabowski J. Usage-based automatic detection of usability smells. In *Human-Centered Software Engineering: 5th IFIP WG 13.2 International Conference, HCSE 2014, Paderborn, Germany, September 16—18, 2014. Proceedings*; Springer, Vol. 5, 2014; pp. 217—234.

30. iMotions. *Facial Expression Analysis*, 2005. https://imotions.com/biosensor/fea-facial-expression-analysis/ (accessed 03, 2022).

31. iMotions. *iMotions Unpack Human Behavior*, 2005. https://imotions.com (accessed 03, 2022).

32. Gabert-Quillen C. A., Bartolini E. E., Abravanel B. T., Sanislow C. A. Ratings for emotion film clips. *Behav. Res. Methods* 2015, *47*, 773—787.

33. Bieber C. Interaction and AI-based emotion recognition for user experience evaluation. In *Applied Research Conference*, 2022.

34. Mahtani K., Spencer E. A. *Hawthorne Effect*, 2017. https://catalogofbias.org/biases/hawthorne-effect/ (accessed 08, 2022).

35. Höfling T. T. A., Küntzler T., Alpers G. W. *Automatic Facial Expression Recognition in Standardized and Non-Standardized Emotional Expressions*, Vol. 13, 2021.

Bionotes

Dominick Leppich
Nuremberg Institute of Technology, OHM-UX, Nuremberg, Germany
dominick.leppich@th-nuernberg.de

Dominick Leppich has been a scientific researcher at the OHM-UX in the Nuremberg Institute of Technology, Germany, since November 2021. He does research on AI-based emotion recognition in user interaction data. In further projects, he acquired advanced knowledge of Virtual Reality (VR) technologies and the simulation of Virtual Prototype's (VP) behavior with state machines.

Carina Bieber
Nuremberg Institute of Technology, OHM-UX, Nuremberg, Germany
bieberca73037@th-nuernberg.de

Carina Bieber has been a graduate student at the Faculty of Electrical Engineering, Precision Engineering and Information Technology (efi) of the Nuremberg Institute of Technology, Germany, since October 2021.

She is a member of the OHM User Experience Center (OHM-UX) and does research on AI-based emotion recognition in user interaction data. She has experience in usability evaluation of user interfaces from fitness devices and user interfaces.

Katrin Proschek
Nuremberg Institute of Technology, OHM-UX, Nuremberg, Germany
katrin.proschek@th-nuernberg.de

Katrin Proschek works as a researcher and UX professional at the OHM-UX in the Nuremberg Institute of Technology, Germany. With an M.A. in educational media and a background in engineering, she is a certified HCD consultant with 25 years of experience to introduce and run HCD processes for the German industry as well as for international research projects.

Patrick Harms
Nuremberg Institute of Technology, OHM-UX, Nuremberg, Germany
patrick.harms@th-nuernberg.de
https://orcid.org/0000-0002-9988-2028

Prof. Dr. Patrick Harms has been Professor of Usability at the Nuremberg Institute of Technology, Germany, since October 1, 2020. He heads the OHM User Experience Center and is a member of the Artificial Intelligence Center of the university. He is an expert in AI-based usability evaluation of user interfaces. In this area, he completed his dissertation in 2015 with a focus on websites and subsequently extended the topic to XR.

Ulf Schubert
DATEV eG, User Experience & Touchpoint Design, Nuremberg, Germany
ulf.schubert@datev.de

Ulf Schubert currently works as Director UX & Touchpoint Experience at DATEV. For many years he has been supporting companies as a manager and consultant in becoming more successful through success-enhancing experiences and attractive product design. He believes in the success of self-organized organizations that learn and improve through interactions with people. His professional focus is on Experience Leadership, Human Centered Organisation, Experience Strategy and Experience/Product Design.