

Research Article

Interaction Modes for Emergency Mobile Apps

Claudia Nass, Jessica Jung, Eduard C. Groen, Karina Villela, and Konstantin Holl 

Fraunhofer Institute for Experimental Software Engineering (IESE), Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

Correspondence should be addressed to Konstantin Holl; Konstantin.Holl@iese.fraunhofer.de

Received 4 July 2017; Accepted 23 October 2017; Published 1 February 2018

Academic Editor: Sergio Mascetti

Copyright © 2018 Claudia Nass et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

While seemingly irrational behaviors such as panicking or displaying antisocial behavior are the responses to emergency situations the media and movies lead us to believe, several studies show that people rather react based on decision-making such as acting altruistically and protectively. However, what can we really expect from people in a crowd in terms of participation in an emergency response system? In this paper, we present a mobile application called the RESCUER App, which allows civilians to participate in the emergency response process by providing information about the emergency to a command center and to receive instructions from this command center. We developed a human reaction model for emergencies to better understand the human-computer interaction capabilities of people in an emergency situation. Based on this model, we defined three different interaction modes: one-click interaction, guided interaction, and chat interaction. These interaction modes were implemented in an interactive prototype and evaluated in an experiment in which high cognitive load was induced to simulate a stress situation, similar to the stress experienced in an emergency. The experiment results showed that the three predefined interaction modes enabled people to interact with the RESCUER App even though they were in a stress situation.

1. Introduction

Every year, the German state of Rhineland-Palatinate organizes a large festival in the region, with more than 200,000 visitors over the course of the three-day event. People can enjoy local food, buy craftwork, or attend concerts by national artists. Now imagine that a market booth in the middle of the festival grounds catches fire. Some festivalgoers close to the booth will call the emergency number as they cannot quickly find members of the event's safety personnel. They will try to describe the place of the incident, but they may not know exactly where they are as the event location is not familiar to them. They will also try to explain what they see around them, which is not easy for them. On the other side of the line, the command center will try to guide the callers through some questions to obtain relevant information from them, while also alerting the security staff and firefighters at the festival venue. The exact site of the incident will still be unknown at this time, and the situation description will be very imprecise. After a few minutes, the number of calls to the emergency number will increase rapidly. In the meantime, the firefighters present at the venue will try to reach the exact location of the

incident and start fighting the fire, but when they reach the fire, it may already have spread rapidly so they will need to call for additional firefighters via the command center. Communication via radio is very short and not persistent, so the command center will still lack vital information about the emergency.

This adverse scenario is exemplary for various problems that may occur in an emergency in a given location with high density of people, that is, that involves a crowd [1]. Typical problems for this scenario include the overloaded communication lines of the command center, which prohibits people from calling for help or providing information about injured people at the site of the incident, and the delay in identifying the exact position of the incident, which may lead to an expansion of the fire and the situation to run out of control. Additionally, the lack of information about injured people, the exact number of people affected (i.e., the people close to the site of the incident who are likely to be affected by the adverse consequences of the fire), or even the physical situation around the fire itself prevent the seamless execution of the emergency response plan and put in danger the success of the operation.

In our work, we try to solve these problems through the use of a mobile application called the RESCUER App, which supports the notification and characterization of an emergency situation by involving the crowd at the place of an incident. Our goal is to stimulate crowd participation just after the incident has occurred when they are still close to the site of the incident, so they can inform the command center about the incident and describe its main characteristics or provide essential details.

The scenario presented above is a typical emergency involving a crowd. An emergency is an event or situation that threatens to inflict serious damage on human welfare (e.g., loss of human life, illness, injury, homelessness, damage to property, or disruption of supplies) or the environment (e.g., contamination of land, water, or air with biological, chemical, or radioactive matter) [2]. Past views assumed that people act out of a reflex in emergencies, thereby showing behaviors that may seem irrational to other persons, such as antisocial behavior or panicking. News broadcasts and the media in general have programmed our minds with this picture [3]. However, recent research suggested that this kind of reaction is a myth without empirical basis. According to Quarantelli [4] and Drury et al. [5], studies and analyses of past emergency situations show that people make decisions during an emergency, for example, the decision of behaving altruistically. They behave calmly and know how to protect themselves and others in their social sphere. However, part of the decision-making process is influenced by the presence of others; in a crowd, people defer responsibility more easily and assume others will act, leading to phenomena like the bystander effect or pluralistic ignorance [6].

The collaborative behavior of people in a crowd involved in an emergency, which was described by Drury et al. [5], is essential for the emergency response process. The use of technologies such as smartphones and other mobile devices increases the resilience of people when responding to an emergency [7]. The point, however, is to know what we can really expect from people in terms of how they interact with a mobile application just after an incident, while they are still in the immediate vicinity of the site of the incident. In particular, we focused our study on the interaction with a mobile application for smartphones, as smartphones are much widely used than other mobile devices such as smart watches. In addition, smartphones allow a richer interaction, as people can take photos and make videos.

Currently, several mobile crowdsourcing-based platforms are available that support people involved in disasters as well as the workforces that respond to such emergencies. The Ushahidi crisis map platform [8] and iLigtas [9] are examples of free general public license software for information collection and visualization, and interactive mapping. Both platforms provide a mobile application that allows sharing incidents via SMS, geotagging, and social media feeding. The platforms create infographics visualized on maps for resource distribution, including medical care, shelter, and food. Both platforms support emergency task coordination through social networking and virtual teaming, geospatial visualization for analysis, and decision-making support.

The platforms InSTEDD [10, 11] and Sahana [12] are open source software platforms that support collaboration and improve the information flow to better deliver critical services to vulnerable populations. Both platforms allow a quick adaptation of the emergency response process and tools to the current emergency scenario, and they additionally offer tools to assist groups in analyzing and visualizing multiple streams of information (SMS, geotagging, and social media) and postdisaster tools for tracking projects and resources.

All these solutions are designed to support long-term interactions especially in the context of natural disasters, where the response process might span days or weeks, and the occurrence of immediate stressors might have already decreased. In contrast, RESCUER (an acronym for “Reliable and Smart Crowdsourcing Solution for Emergency and Crisis Management”) focuses on the immediate process running concurrently with the identification of an incident. Specifically, the RESCUER project aims at developing a complete platform to support command centers in quickly handling emergencies and managing crises based on the reliable and intelligent analysis of crowdsourcing information [13]. The application scenarios it focuses on are incidents during large-scale events and in industrial areas. The RESCUER App is one the main components of the RESCUER platform, which also includes Data Analysis Solutions, the Emergency Response Toolkit, and the Ad-hoc Communication Solution.

For the design of the RESCUER App, we modeled human reactions in emergency situations and derived three interaction modes that support user interaction with a mobile device in such circumstances: one-click interaction, guided interaction, and chat interaction. These interaction modes provided concrete requirements for the user interface that were implemented in the RESCUER App accordingly. During an earlier stage of the project, we performed an experiment with a prototype of the RESCUER App in which stress was induced in participants to simulate that they found themselves in an emergency situation. The results indicate that the three predefined interaction modes enable people to interact with the RESCUER App even when they are in a stressful situation.

In Section 2, we present the related work that supported the definition of the RESCUER Human Reaction Model. The model itself is presented in Section 3. Based on that model, we derived the requirements for the interaction design of the RESCUER App, which led to the three concrete interaction modes described in Section 4. Section 5 presents the resulting RESCUER App. In Section 6, we describe the experiment that we conducted, and in Section 7, we discuss how it confirmed that the three predefined interaction modes support the interaction of the crowd with the mobile application even in a stressful situation. Finally, we discuss the main outcomes in Section 8. Conclusions and future work are addressed in Section 9.

2. Related Work

In this section, we present relevant work and definitions that formed the basis for the specification of the human reaction

model used for the development of the RESCUER App. Pan [14] presents a human and social behavioral model for typical emergency situations in large-scale event scenarios that involve overcrowding and crushing, which among others occur at sports stadiums, schools, or social gathering places such as nightclubs. His goal was to build a multiagent simulation system for egress analysis. He identified five typical evacuation behaviors, which were modeled in a computational framework: queuing, competitive, leader-following, altruistic, and herding behavior. These behaviors are related to the output of the decision-making process that is responsible for identifying and properly responding to an emergency. Factors influencing this process and thereby influencing human behavior include

- (i) human physical characteristics (i.e., body size, mobility, age, and gender),
- (ii) environmental characteristics (i.e., geometric constraints, emergency type, and egress systems),
- (iii) psychological and sociological characteristics (i.e., individual, interaction among individuals, and group).

Gannt and Gannt [15] propose that in order to make emergency plans and processes in industrial area scenarios more efficient, they should also be based on typical human cognitive capabilities and behaviors. They argue that most emergency plans and systems ignore such human capabilities and, consequently, expose the employees and the environment to higher risks. They paid special attention to the cognitive decision-making process that people execute during emergencies, the influence of the changes in the environment, and especially the social behavior that emerges during an emergency. The authors argue that social bonds between people involved in an emergency are solidified and even newly created.

Similarly, Perry and Greene [16] describe a decision-making process that is executed by individuals who identify and properly respond to an emergency. They characterize human behavior in a disaster as “non-traditional behaviour in response to a changing or changed environment” [16]. This means that individuals who face unexpected environmental changes reexamine their own behavior in relation to the altered conditions and adapt their behavior in order to protect themselves and minimize harmful consequences. This process of behavior adaptation consists of three phases: risk identification, risk assessment, and risk reduction. In the case of an emergency, the changes in one’s immediate environment are recognized as a threat to the individual, and the individual takes protective actions accordingly.

According to Leach [17], the decision-making process is managed in the human working memory as a neurocognitive function. The processing of such information involves several steps in order to turn the perception of a threat into an appropriate action. Under optimal conditions, this takes a minimum of eight to ten steps, but this number can increase due to higher task complexity and physiological factors that influence the individual. The processing of an emergency often results in “freeze” behavior

experienced by the individual just after he or she has perceived the threat. Leach [17] classified three typical behaviors in response to threats:

- (i) The first group (10–15% of the people involved in an emergency) will remain calm, think quickly, and retain situational awareness, hardly affecting their judgment and reasoning abilities.
- (ii) The second group (about 75% of the people involved) will be stunned and show impaired reasoning and sluggish thinking.
- (iii) The third group (10–15% of the people involved) will show confusion, paralyzing anxiety, and exhibit counterproductive behavior that adds to their risk.

The term “stress” is frequently used in the context of emergencies such as disasters. However, stress is an ambiguous term, as it can have different meanings. Broadly speaking, stress is the biophysiological reaction to a stressor, which in the case of an emergency is the threat. Stressors (i.e., that what causes stress) are “circumstances that threaten a major goal, including the maintenance of one’s physical integrity or one’s psychological well-being” [18]. The psychological response to a stressor is called distress, which can manifest itself through affective states, such as anxiety, frustration, euphoria, or sadness.

Kemeny [18] proposes an integrated specificity model of stress. In this model, the exposure to a specific stressor evokes an integrated psychophysiological response (which includes emotion, motivation, and physiological responses) that intends to generate a protective action in an individual in reaction to the threat. In this model, the resources available for reacting to a threat moderate the relationship between a stressor and its psychophysiological response. This means that when faced with a threat, the level of stress is regulated by one’s own motivation, the physiological responses to the threat, and the individual’s resources for reacting. Resources, in this case, may include time, tools, abilities, and space.

Finally, Staal [19] explores the interaction between stress, cognition, and human performance. He found that various stressors can affect cognition in which case these stressors contribute to a reduction in the working memory’s performance, especially in combination with a low self-efficacy to averting a threat. Under stress, the resources available for the working memory are reduced, in part due to the physiological changes related to the fight-or-flight response. Considering that when faced with a threat in an emergency, the decision-making process takes place in working memory, stress can consequently affect the decision-making process and its execution. Thus, the level of stress experienced by a person in the face of a threat can diminish his or her cognitive capabilities and performance.

3. RESCUER Human Reaction Model

The RESCUER Human Reaction Model is based on the models presented above and uses the ABC Model of Attitudes [20] as a basis for distinguishing between various

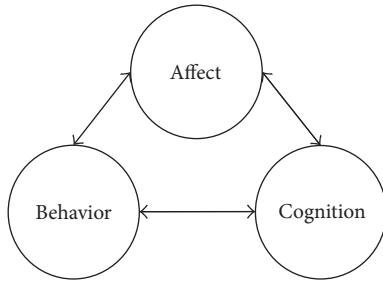


FIGURE 1: Interplay between affect, behavior, and cognition.

human manifestations. The model is a simplified representation of what happens within the human organism, separated into three components: affect, behavior, and cognition. The affect component describes the emotions that individuals experience during an emergency, that is, what people feel. The behavior component presents different possible behaviors during an emergency, that is, what people do. The cognitive component presents the decision-making process that people execute during an emergency, that is, what people think. As Figure 1 illustrates, these three components bilaterally influence each other.

The result of this interplay characterizes the human reactions in the face of an emergency on a high level of abstraction. In practice, the three components are not easily distinguishable because humans are unified organisms with complex internal interrelationships. However, separating what is essentially a united system into three integrated components helps us approach human reactions in emergencies from multiple points of view and beyond the “black-box” approach in which behavior is the only measurable and perceivable output.

Figure 2 shows the RESCUER Human Reaction Model. Each of the components will be described in the following sections. We will start with the cognitive component, as this component is responsible for the basal thinking process [20]. For a better understanding of Figure 2, we also recommend reading the model starting from the cognitive component at the bottom of the picture.

3.1. Cognitive Component. The cognitive component represents the decision-making process of people in an emergency situation [16]. This process takes place in the human working memory [17] and can be subdivided into three phases: (1) risk identification, (2) risk assessment, and (3) risk reduction.

During risk identification, people assess whether there is a real threat. Factors that influence this decision are the credibility of the authority that is issuing a warning, environmental cues, and the reaction of the social group. In large-scale event scenarios, information coming from authorities (e.g., police or security personnel) is interpreted as being more trustworthy than information from civilians. Environmental cues are a stimulus for gathering information to select the appropriate response. For example, when hearing a fire alarm, people tend to look for a sign of fire, the smell of smoke, or other perceivable indicators of a risk. The third factor is the social group. People look for confirmation

that there is a threat by observing the behavior of other people nearby who would theoretically also be affected by the threat. If other people have identified a threat and react accordingly, this will increase people’s personal confidence in the severity of the threat. For example, in the case of a fire alarm at a company, employees will observe their colleagues’ response. Do they leave the workplace in the direction of a meeting point, or is it a false alarm? If there are inconsistencies between the three factors, people become uncertain and may remain in a closed loop of risk identification until the inconsistencies are resolved. In this state of uncertainty, people tend to follow others until they are able to determine if there is a risk for themselves.

If people recognize that the threat is real, they initiate a risk assessment to determine their individual risk [16]. This involves analyzing the content of the emergency warning, past experiences, and their own knowledge about the threat and its expected consequences. The issued emergency messages should provide information about the probability of impact, the proximity of such an impact, and the severity of the impact. The more detailed the emergency warning, the higher its credibility, whereas the more inconsistent or unclear the message, the harder it is for people to assess the real risk [16]. Past experiences with emergency situations may influence the overall risk assessment positively or negatively. For example, people who have frequently experienced false alarms tend to take alarms less seriously. The third factor influencing the risk assessment of an emergency situation is previous knowledge about a specific threat. For example, if people are well prepared for abandoning the building in the case of a fire alarm and know what the available resources are for reacting to a fire, they will assess the situation to be less severe compared to an emergency they have no knowledge about or for which they have no resources.

Based on the severity level of the emergency, people determine the effectiveness and their chances of responding to a threat in a particular way. If they think that they stand a good chance, they will contemplate the possible action scenarios to reduce their personal risk and/or minimize the impact of the threat and choose one plan. This could, for example, be an official plan, that is, a plan recommended by authorities in charge of security and safety. Once a plan is selected, the feasibility of its execution is considered against factors such as the available time and other resources. People then display the behavior that corresponds to the plan they have in mind. People who decide that a reaction is impossible or insufficient do not take any protective action against the threat at a cognitive level.

3.2. Behavior Component. This component describes how people act during an emergency. Environmental characteristics influence the action plan chosen by the people involved in an emergency. Table 1 presents typical behaviors in emergencies [14].

Figure 2 presents typical relations between the decision-making process in the cognitive component and the resulting behavior during an emergency with the evacuation of people in large-scale event scenarios. Particularly when

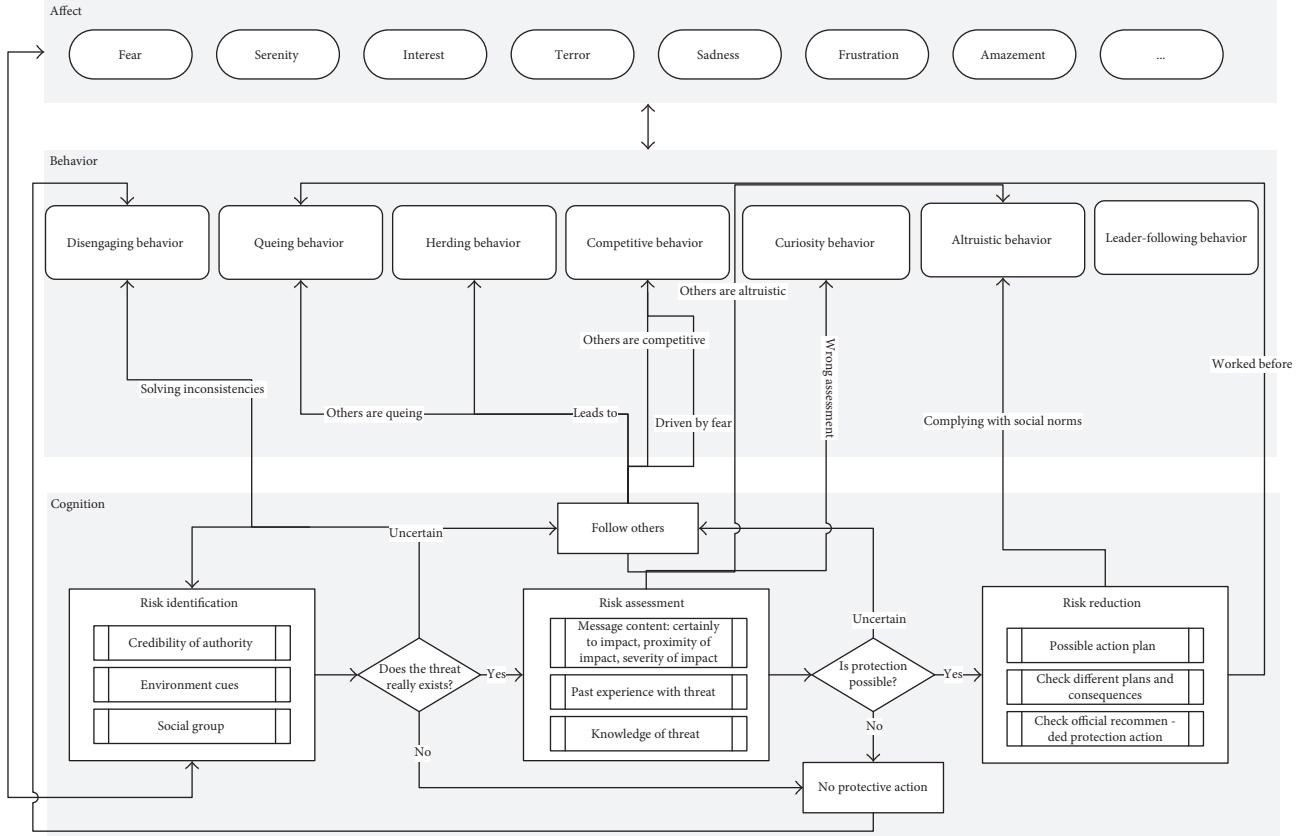


FIGURE 2: RESCUER Human Reaction Model.

TABLE 1: Typical behaviors in emergencies.

Behavior name	Behavior description
Disengaging behavior	Individuals are withdrawn and inactive. They do not look for an exit, adding to their both own risk and that of others, as they block exits and escape ways and interrupt the evacuation flow.
Queuing behavior	People self-organize queues to get to the exits; the crowd streams out in an orderly way. Queuing behavior does not lead to blockages, but to a more effective evacuation.
Herding behavior	People evacuate a room through one obstructed exit, while other exits are not fully utilized. This is typical behavior of individuals who are uncertain about what to do and decide to follow others.
Competitive behavior	Individuals compete to exit, violate the personal space of others, and do not respect social structures and rules.
Curiosity behavior	People move closer to the threat, do not take any protective action, take photos and discuss the situation with other curious individuals. This phenomenon is known as “disaster tourism.”
Altruistic behavior	People help other individuals during the evacuation, especially when the other person is impaired. People observing the willingness of others to help also tend more towards helping. People believe they are following social norms or are being ethical by helping others in need.
Leader-following behavior	People tend to follow others of a higher hierarchical rank during an evacuation. Sometimes a leader emerges when the situation is uncertain to many people.

people should leave the site of an incident but encounter inconsistencies while identifying the risk, they may tend to disengage and fail to react to the threat until the inconsistencies are resolved. In the case of high uncertainty, people tend to follow others, exhibiting queuing, herding, or competitive behavior. If people decide to take protective actions, they will most likely follow an assumed leader, for example, a family member who is deemed to be more rational or an officer of an operational force.

3.3. Affect Component. The affect component refers to a person's feelings and the intensity of those feelings, that is, the emotions triggered during an emergency in response to physiological and cognitive factors [21]. Plutchik [21] describes eight basic human emotions, consisting of four opposite pairs at different degrees of intensity, which are presented in Figure 3.

Negative emotional states are strongly associated with emergencies involving stress. Feelings such as fear, anxiety,

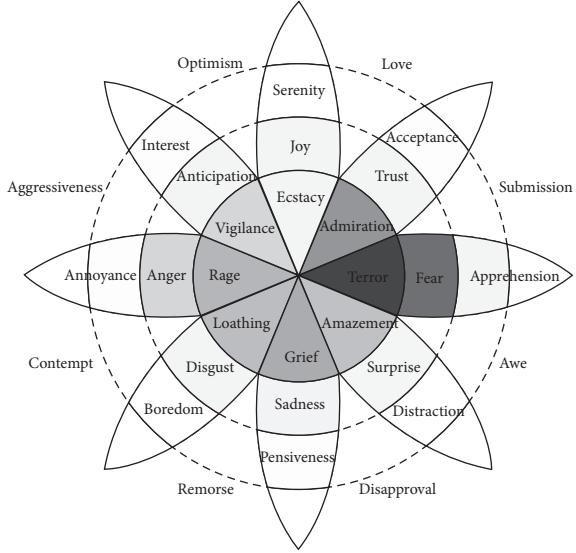


FIGURE 3: Plutchik's wheel of emotional states [21].

irritability, embarrassment, depression, helplessness, euphoria, frustration, and hostility are the most common feelings associated with emergencies [20]. Some of these typical feelings are associated with certain behaviors, but the direct correlation between specific feelings, behavior, and cognition is very complex and difficult to characterize in detail.

Emotions are among the most obvious (and best visible) aspects of human reactions. Humans are capable of perceiving the physiological response of others, such as variations in heart rate, respiration, and facial expressions, and thereby interpret their emotional state. This, in turn, may influence their own emotions. This is why the human reaction in emergencies is often described in terms of affect, and why the emotions experienced and observed in a crowd form an important stimulus for individuals in determining how to relate to the situation [22]. For example, if people in the crowd feel anxiety, they show it through their expressions and behavior, which may influence others who perceive it.

3.4. Reaction Changes, Stress Factors, and Interaction Capabilities. Even when an individual has developed a certain reaction in response to an emergency, this reaction can still be subject to change. The cognitive decision-making process is continuous and capable of adapting quickly. New or changed environmental stimuli received are used to reassess the severity of the situation and adapt the action plan accordingly.

Stress is one of the key factors affecting human reactions in an emergency situation. According to Staal [19], the definition of stress as the interaction between the following three elements includes the contemporary assumptions about the concept of stress:

- Perceived demand is the requirement for executing the plan elaborated by the individual to protect him/herself and to minimize the threat.

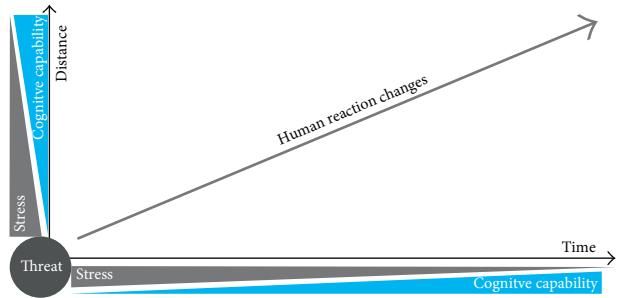


FIGURE 4: Changes in stress, human reaction, and cognitive capabilities.

- Perceived ability to cope with the demand is the subjective assessment of the resources the individual has available to execute the plan.
- Perception of the importance of being able to cope with the demand is the affect-based motivation of the individual to execute the plan and achieve goals.

This definition implies that as long as an individual asserts that he or she can respond to the stressor using the currently available resources and has the internal motivation to do so, the stress level will be low. In contrast, if the resources or the motivation cannot fulfill the demand, the perceived self-efficacy will be low, causing the stress level to be high. In this context, the distance and time from the threat affect the three elements that describe stress. The closer an individual is to the threat, the higher his or her level of stress will be. Similarly, right after the perception of a threat, the level of stress will be heightened by the individual's state of uncertainty and anxiety. Figure 4 illustrates this influence of distance and time on the level of stress.

- The time span between the occurrence of the threat that triggered the emergency and the first moment the person is confronted with the situation is a factor that influences the stress level. The more recent the threat occurred, the less information about the emergency is available and the fewer actions have been taken to control the situation. This causes stress, because people at the site of the incident causing a threat have not had enough time and information to assess the threat and they are primarily concerned with protecting themselves and possibly others.
- The distance between the exact place of the incident causing a threat and the person has a direct impact on the stress level. The closer the victim's position to the epicenter or danger zone of the threat, the higher the physical threat (e.g., heat, noise, and water).

Because stress impairs decision-making [19], fewer stimuli are attended to. Particularly affected are the encoding, rehearsal, and retrieval of information. This means that individuals are less capable of understanding and recognizing information when under stress, especially the closer they are to a threat in terms of time and distance. When the stressor is of an emotional nature (e.g., related to a state of anxiety), peripheral events are remembered even

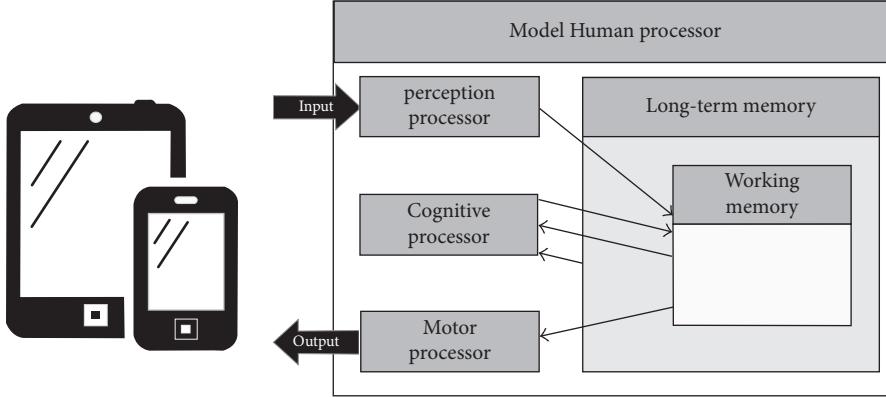


FIGURE 5: Model human processor adapted from Card et al. [23].

less often [19] because individuals tend to use as many mentally available resources as possible on solving the primary task, which is the decision-making process. The reduced attention for peripheral events affects the perception and processing of the perceived stimuli, causing changes in the environment that are not directly related to the primary task to be overlooked or to be registered without processing.

The effects of stress on the cognitive capabilities of a person have major implications on the interaction concept for software systems designed to interact with persons in an emergency situation. Interacting with a software system during an emergency situation requires focusing one's attention and dedicating resources from the working memory to the interaction, especially if the system asks a person to report on the perceived environmental cues and characteristics. Figure 5 presents a model human processor, which reflects how the processing of information perceived during the interaction with a software system occurs in working memory [23]. As part of the interaction with the software system, not just the perception plays a role but also the way of providing input, for example, by operating a touch screen. This is why the preparation of the motor processor to respond to the human-computer interaction stimuli is also required [24] and included in the model human processor accordingly. Breiner [25] describes this interaction process as follows:

All information that is perceived by a human (stimulus, e.g., visual, auditory, tactile) acts as input for the perception processor. After the execution phase of this processor, the information is available in the working memory. Based on the long-term memory, the information of the working memory is processed by the cognitive processor. The result of this process is again stored in the working memory. The outcome of the human computation in the working memory then triggers the motor processor to execute the desired action (e.g., press a button, make a statement, etc.). Elements of the working memory are linked to the corresponding elements in the long-term memory. Based on previous experience, data and actions that will be executed depending on the content of the working memory are stored in the long-term memory.

The stress level may impair people in their interactions with software systems—in our case, a mobile application—because the available resources are being used for the decision-making process. Interaction with mobile devices therefore requires the stress level to be sufficiently reduced to have enough information processing resources available for the interaction.

Additionally, several authors [19] have demonstrated that well-learned tasks tend to be more resistant to the effects of stress. The practice and frequent activation of an operation tend to commit this operation to the long-term memory, which is less susceptible to being influenced by stress. The activation of these memories would allow the user to perform actions with a kind of automaticity, for which fewer mental resources are required during planning and execution [19]. Hence, if the required interaction in an emergency situation is practiced when no emergency is taking place, this may make the interaction easier, or even make it possible for an interaction to take place that would not have been possible without practice.

4. Interaction Modes and Requirements

Considering the RESCUER Human Reaction Model presented in Section 3, concrete requirements for the interaction design of the RESCUER App were defined in order to support the participation of the crowd in the response process of an emergency. As the level of stress in an emergency has an impact on people's ability to interact with mobile devices, our solution strategy provides an interaction spectrum for information gathering with three interaction modes: one-click interaction, guided interaction, and chat interaction. According to our theory, the closer the distance to the exact place of the incident and the more recent the incident, the more information has to be gathered automatically, because the cognitive capabilities of the people present will still be limited so that they cannot provide all the information a command center requires for the emergency response process. Figure 6 depicts this relationship between the available cognitive capabilities and the level of stress in an emergency, indicating when each of the aforementioned interaction modes is most appropriate.

The current characterization of the three modes is as follows:

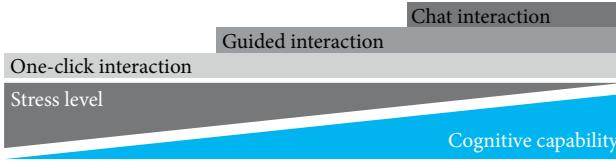


FIGURE 6: Stress, cognitive capabilities, and interaction modes.

- (i) *One-click interaction*: Since people's cognitive capability is mainly consumed by their decision-making processes in response to the emergency situation itself, they are unable to respond to any request or complete any task through the user interface. Therefore, any mobile application to be used in an emergency situation has to be able to automatically gather all the information about the situation that is relevant for the workforces.
- (ii) *Guided interaction*: When people are still not able to respond to complicated interaction demands but can at least provide a little information about the emergency, a guided approach ensures that the interaction is kept as simple as possible. These people are likely still near the site of the incident and are too distracted to execute a large set of actions with their mobile devices. Therefore, it is advisable to pose precise questions (e.g., by using multiple-choice selections) and ask them to execute easy and fast actions (such as sending them to specific areas of the venue).
- (iii) *Chat interaction*: When eyewitnesses of an emergency are safe, they may become observers and can contribute information about the situation more intensively and actively, and at a greater level of detail. Their cognitive load is no longer fully consumed by the situation, and their ability to interact with their mobile devices the way they usually do has increased. Due to the physical distance and the elapsed time, the information that can be gathered from the physical sensors of the mobile device might not be as important as in the other modes.

As people using an interactive system in an emergency situation cannot be pressed into a concrete interaction mode as it is difficult to know their specific level of stress, the proposed interaction spectrum spans an escalation order of the three interaction modes. Interaction through in-air gestures and device shaking are not part of any proposed interaction mode. First, they would require users of the mobile application to previously learn the expected movements. Second, the user must be protected from entering wrong input or input that may lead to dramatic consequences. When the level of stress is high and the cognitive capabilities are consequently low, the use of these interaction means is too error-prone. On the other side, when the level of stress reduces and thus the cognitive capabilities increases, users are expected to provide more information than is possible through in-air gestures and device shaking.

TABLE 2: Requirements of the interaction modes.

Interaction mode	UI requirements
One-click interaction	As users are highly impacted by stress or cognitive overload, a single click is the most complex interaction that can be expected. This can be taken as a confirmation of intent, resulting in fully automated information gathering. This means that all mechanisms responsible for gathering data can be activated.
Guided interaction	As users are not directly impacted by the incident anymore and cognitive workload is available again, more focused interaction can be expected (but not too complex). This can be implemented through simple multiple-choice selection, where the user is guided to a possible answer by being confronted with various answers. By recognizing the correct (or most suitable) answer, the user is able to provide more precise information about the incident.
Chat interaction	Unstructured/open information. Having almost their complete cognitive power available, users can execute more complicated tasks. This means reacting to individually defined tasks or providing a description in plain text. Here, basically all interaction possibilities can be integrated into the app.

Table 2 presents concrete user interface requirements that were elaborated on the basis of the different interaction modes introduced above.

5. The RESCUER App

Figure 7 illustrates a prototypical implementation of the three interaction modes in the RESCUER App. We assert that, as long as people have enough cognitive capabilities, they will interact with the application and move from one interaction mode to the next. If the working memory can no longer support the interaction with the RESCUER App, people will naturally stop using it and focus on the main decision-making task, for example, fleeing from the site of the incident.

Figure 8 presents the navigation in the case of a fire emergency. When touching the fire icon (orange), the user is taken to a confirmation screen and subsequently to a screen where he or she can describe the characteristics of the fire. It is also possible to take pictures and submit the report. Next, the chat screen appears, where the user can provide additional information as well as send more photos and videos to the command center. This means that the interaction modes appear in the same order as the model describes: the easiest screen first, that is, one-click interaction, followed by the more complex screens, namely guided interaction and chat interaction, which is the last screen.

In the one-click interaction mode, people have to select which type of emergency they want to report. They are presented with a set of predefined options of typical incident types (fire, explosion, gas leakage, and environmental emergency) in the application context of the RESCUER App. In addition to the emergency type, the RESCUER App

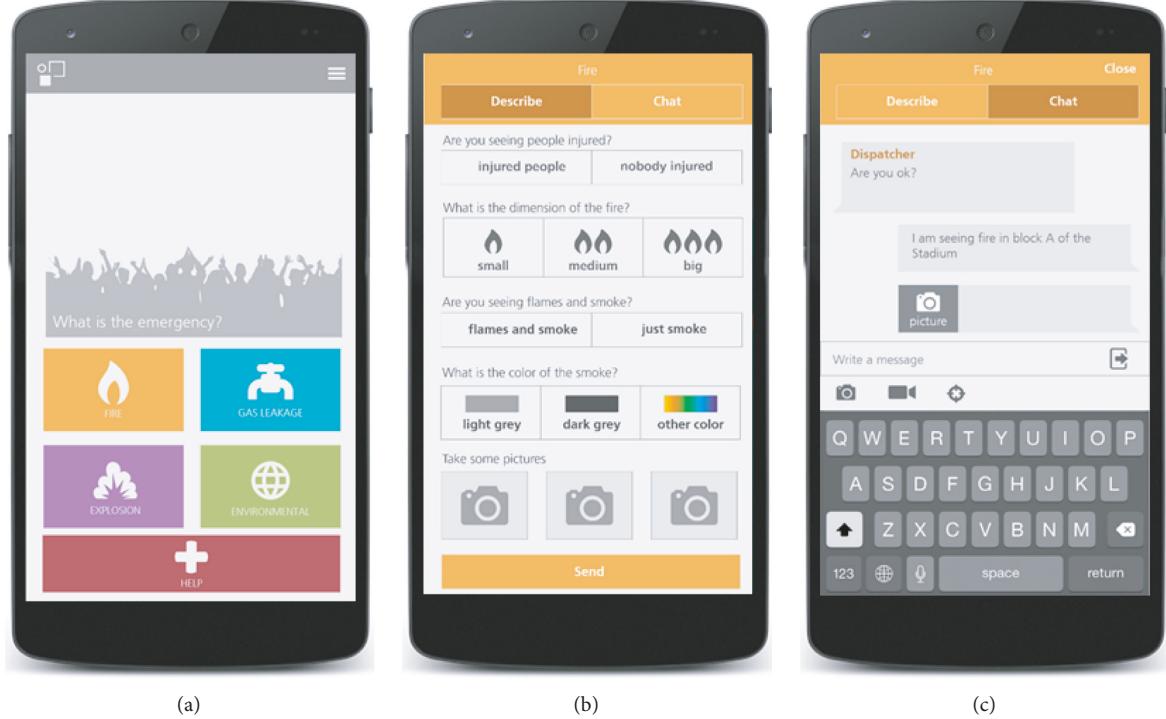


FIGURE 7: The RESCUER App and interaction modes. (a) One-click interaction. (b) Guided interaction. (c) Chat interaction.

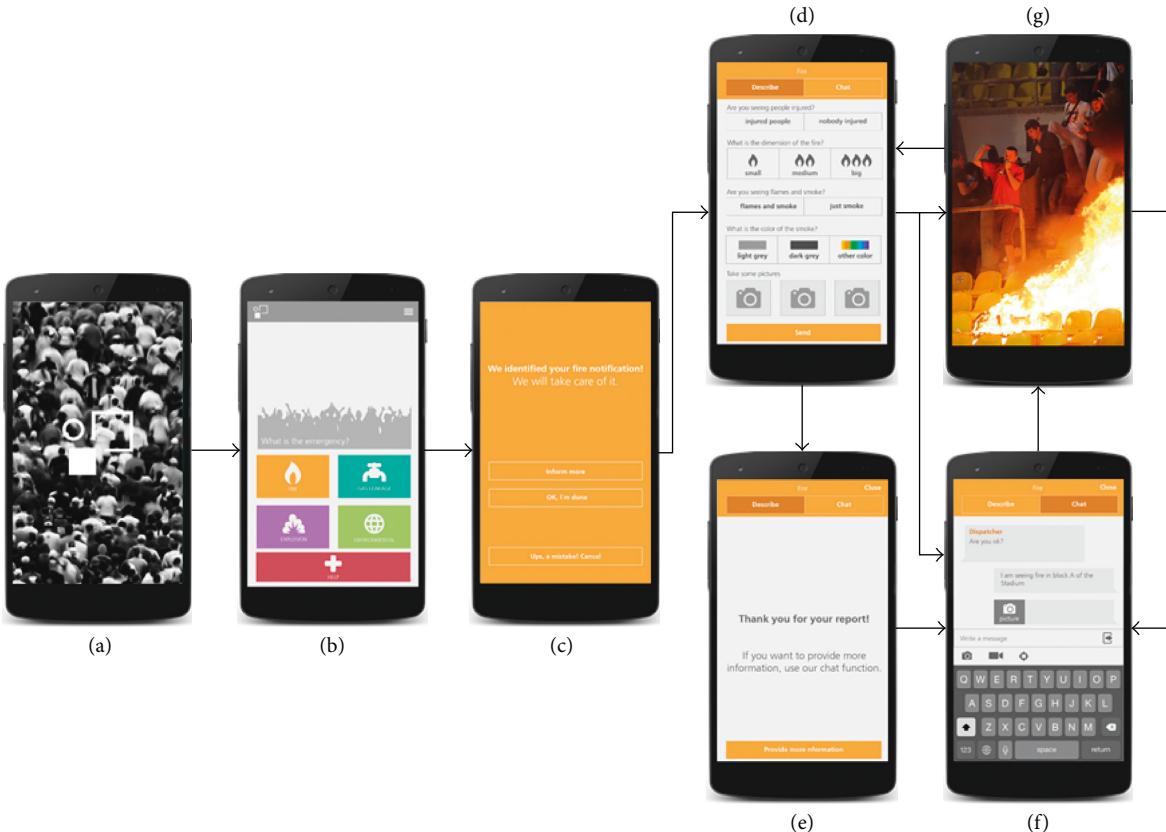


FIGURE 8: Navigation of the RESCUER App through the interaction modes.

automatically provides information about the position of the person sending the report and the time at which the report was sent. Following this one click, the RESCUER App can automatically collect relevant data from users located close to the site of the incident with the help of the physical sensors of the users' mobile devices.

In the guided interaction mode, people answer typical questions that characterize the emergency situation by reporting what they see and send pictures. During the development of the RESCUER App, workforce representatives have helped to select the most relevant questions to be answered by either the general crowd or the workforces at the site of the incident (see screen (b) in Figure 7).

The chat interaction mode screen is a typical chat functionality as encountered in typical mobile communication tools. It allows people to write free text, dictate text, and send pictures and videos to provide information on their own initiative, or in response to specific questions asked by the command center.

6. Experiment

To validate the defined interaction modes and to identify improvement potential for the RESCUER App, we performed an experiment in a laboratory environment with 26 participants. The aim was to explore different aspects of the theory that supports the definition of the interaction modes as well as to observe how people interact with the design conceived for the RESCUER App when exposed to different degrees of stress. Four main research questions guided the experiment with the RESCUER App.

6.1. Research Questions

RQ1. Mental effort increases from one mode to the next in the following order: one-click interaction, guided interaction, and chat interaction. This increase becomes more intense under high cognitive load conditions.

We assumed that users would perceive low mental effort when interacting with the screen representing the one-click interaction, that the perceived mental effort would increase to a medium level when interacting with the screen representing the guided interaction mode, and finally, that people would perceive high mental effort when interacting with the screen representing the chat interaction mode. This increasing perception of mental effort was expected to be further intensified during the execution of an additional task they must perform in parallel, that is, when participants experience a higher cognitive load. In this research question, we have two independent variables, namely, interaction mode and cognitive load. The dependent variable, which we measured with the help of the Subjective Mental Effort Questionnaire (SMEQ) [26], was the perceived mental effort. The analysis method we used was a repeated measures ANOVA with two within-subject factors.

RQ2. Previous involvement with an emergency situation does not affect the perceived mental effort when using the RESCUER App under different levels of cognitive load.

The aim of the RESCUER App's design was to develop a mobile application that is easy to use in cognitively demanding situations regardless of a person's previous experience with emergencies. We assumed that having previous emergency experience results in no differences in the SMEQ ratings for the different interaction modes and the different levels of cognitive load, which means that it should be possible to use the RESCUER App equally comfortably, regardless of previous experience with emergencies. To analyze this research question, we used the repeated measures ANOVA model from RQ1 and included a third independent variable relating to previous emergency experience with two conditions (no previous experience versus previous experience) as a between-subject factor. The dependent variable is the perceived mental effort (SMEQ) as in RQ1.

RQ3. The duration of the interaction with the RESCUER App increases from one mode to the next in the following order: one-click interaction, guided interaction, and chat interaction. This increase in duration is larger in high cognitive load conditions.

With each new screen in the RESCUER App, additional interaction possibilities arise. Therefore, we assumed that the time spent on each screen would increase with the added functionalities. We expected the shortest time span for screen 1 (one-click interaction), a medium time span for screen 2 (guided interaction), and the longest time span for screen 3 (chat interaction). Furthermore, if asked to execute an additional parallel task (an *n*-back task as presented in Section 6.2) while using the mobile application, participants would need even more time to interact with the screens in the three interaction modes. Again, we used interaction mode and cognitive load as independent variables. As the dependent variable, we measured the time in seconds spent on each screen. This results in a repeated measures ANOVA with two within-subject factors.

RQ4. Even under high cognitive load conditions, participants are able to use the mobile application and provide correct answers for a parallel task, the *n*-back task.

The design of the RESCUER App is optimized for usage in cognitively demanding and stressful situations. Consequently, it should be no problem to use the app and still react to external stimuli. As external stimuli, we used the *n*-back task, specifically as one-back task (see Section 6.2 for a description). We assumed that when using the RESCUER App, it is still possible to provide mainly correct answers in the one-back task. In this research question, we investigated the participants' answers in the one-back task, that is, correct answers, incorrect answers, and omitted answers in response to the stimuli (used as dependent variables). We assume that incorrect answers and omitted answers do not differ significantly from zero. Therefore, we use the one-sample *t*-test statistic.

In addition to the above research questions, we wanted to identify existing usability problems and opportunities for improving the usability and user experience (UX) of the

TABLE 3: Experiment design with interaction mode (factor 1) and cognitive load (factor 2).

		Factor 1: interaction mode		
		(A) One-click interaction	(B) Guided interaction	(C) Chat interaction
Factor 2: cognitive load	No/low cognitive load (task 1)	X_{11}	X_{12}	X_{13}
	High cognitive load (task 2)	X_{21}	X_{22}	X_{23}

RESCUER App. Especially in the context of an emergency situation, it is very important that the RESCUER App is without usability problems so that a fast and precise interaction of the crowd with the command center is guaranteed. Therefore, we performed a descriptive analysis of items from a usability questionnaire and implemented one-sample t -tests to check for significance.

6.2. Study Design and Independent Variables. The study design comprised a 2×3 within-subject design, that is, each participant used the RESCUER App twice to report an emergency under different levels of stress caused by cognitive load: the first time with no additional parallel task, resulting in a no/low cognitive load condition (task 1), and the second time with high cognitive load imposed by a parallel task (task 2). Table 3 shows the experimental combination of the independent variables *interaction mode* (factor 1) presented by the RESCUER App in three modes (one-click interaction versus guided interaction versus chat interaction) and *cognitive load* (factor 2) with two levels (no/low cognitive load versus high cognitive load). All participants did the experiment under all experimental conditions.

To induce a higher cognitive load in task 2, we used an audio-based n -back task. An n -back task is a memory task that requires the participant to remember whether the presented stimulus is the same as a stimulus presented n iterations previously [27]. The n -back task traditionally contains spoken word stimuli, but such stimuli could interfere with the language processing required for operating the mobile application [28]. In our study, the auditory stimuli were tones presented for 1000 ms, followed by a 2000 ms interval before the next tone. N -back approaches usually assume 3000 ms trials, consisting of a stimulus and an interstimulus interval. These trials may vary in setup; some use a 1500 ms stimulus with a 1500 ms interstimulus interval (e.g., Schuhfried [29]), while others use a 500 ms stimulus with a 2500 ms interstimulus interval (e.g., Beavon [27]). The stimuli in our study were three sonorous signals in the chromatic scale, with sine wave tones for “low” (261.63 Hz; middle C), “middle” (440 Hz; concert A), and “high” (698.46 Hz, F), generated and sequenced in Audacity (a free audio editor and recorder) [30]. The tones were placed in a randomized order according to a numeric sequence (1 for “low,” 2 for “middle,” and 3 for “high”) generated at <http://www.random.org>, which was modified to repeat the same tone three times at most. During the execution of the experiment, the tones were played as 320 kbps MP3 files from the moderator’s computer. In our experiment, the participants had to perform a one-back task, meaning they had to remember whether the last tone played



FIGURE 9: Stimulus 1 used for the description of the incident (photo taken from Turner [31]).



FIGURE 10: Stimulus 2 used for the description of the incident (photo taken from Wikimedia [32]).

was the same as the tone played right before it. If the tones were identical, the participants were to say “yes,” and if they differed, they were to say “no.”

6.3. Materials. To support the variation of factor 2, two pictures were prepared to support the user in describing a fire situation in a soccer stadium (Figures 9 and 10). The order of the pictures was balanced across the participants over task 1 (no/low cognitive load) and task 2 (high cognitive load). This means that half of the participants were presented with the picture in Figure 9 when performing task 1 and the picture in Figure 10 when performing task 2, and vice versa for the other half of the participants. By balancing the fire pictures, we wanted to reduce the effects that the specific style and event of each picture might cause in the description of the incident when using the RESCUER App.

Because the experiment was conducted in an early phase of the project’s development process, we implemented an interactive prototype [33] of the RESCUER App using the prototyping tool Axure Pro 7. The participants were able to navigate between the screens, answer the questions of the guided interaction mode, and type information in the chat

interaction mode. The camera function was simulated. The participants had to simulate taking a picture during the study, that is, by pointing the camera at the picture of the event and clicking on the trigger on the screen, following which a sample picture was loaded into the RESCUER App.

6.4. Measurement of the Dependent Variables. Several resources were prepared for the measurement of the dependent variables. In this section, we describe the objective and subjective methods used to investigate our research questions.

For the characterization of the participants, a demographic questionnaire was prepared with questions related to their age, gender, highest level of education, musical skills, and previous experience with emergencies (in terms of whether they have previously been involved in an emergency and how often they have called the emergency number). The participants were also asked to describe positive and negative aspects of their previous experience with making an emergency call.

For measuring the mental effort, we used the Subjective Mental Effort Questionnaire (SMEQ) [26]. The SMEQ is a rating scale ranging from 0 (*not at all hard to do*) to 220 (*tremendously hard to do*). Nine textual items are distributed on a vertical scale at irregular distances. Schuette [34] describes the SMEQ scale as having a good implementing cost for the measurement of subjective mental effort as an indicator for cognitive load. Immediately after using the RESCUER App (in task 1 or task 2), the participants were asked to mark on the vertical line the perceived difficulty of the overall task and of the interaction modes (*one-click interaction*, *guided interaction*, and *chat interaction*). Scoring the paper version of the SMEQ during analysis requires measuring the distance in millimeters from the bottom of the vertical line to the marking made by the participants. The millimeter measurements are converted back to the rating scale of the SMEQ, that is, from 0 to 220.

The eye-tracking system ETG 18-1316-361 of Senso-Motoric Instruments (SMI) was used for recording objective data. Eye-tracking systems are used to analyze the eye movements of individuals when they are interacting with a system [35]. Eye tracking allows measuring the degree of interest in certain areas of a user interface, the sequence of fixations on different areas of the user interface, as well as revisits or “stickiness” in certain areas [35]. Pupillary dilations and contractions can also be recorded using SMI’s eye-tracking system. In our study, the eye-tracking system helped to precisely measure the time spent in each of the interaction modes in both tasks, namely no/low cognitive load (task 1) and high cognitive load (task 2).

In order to ensure that the participants were executing the *n*-back task correctly, an observer logged the answers provided by the participants when performing the *n*-back task. In order to do so, the observer checked the answers given by the participant against a template with the correct answers and marked the correct and wrong answers as well as omitted answers on the template.

Finally, to measure the perceived usability, a selection of items of the ISONORM 9241/10 questionnaire were used. The ISONORM 9241/10 questionnaire deals with general ergonomic principles that apply to the design of dialogs between humans and information systems [36]. It considers seven aspects of software ergonomics: *suitability for the task*, *suitability for learning*, *suitability for individualization*, *conformity with user expectations*, *self-descriptiveness*, *controllability*, and *error tolerance*. In the ISONORM 9241/10 questionnaire, each aspect of software ergonomics is explored by several questions. In our study, we selected the most relevant aspects for the RESCUER App and also selected the most relevant questions of each selected aspect in order to have a reasonable balance between the time required for answering the questionnaire and the time spent on using the mobile application itself. The participants assessed nine items from the following selected aspects: *suitability for the task*, *suitability for learning*, *conformity with user expectations*, and *self-descriptiveness*. One statement of the item pair was presented to the participants, who could assess this statement by using a 7-point scale from 1 (*does not apply at all*) to 7 (*fully applies*).

6.5. Experimental Setup. The study was performed over the course of three days in a laboratory-like environment. Each experimental run lasted between 30 minutes and one hour. No run exceeded one hour. The participants were invited individually into a meeting room with a table holding the experiment material, a large display with the pictures of the incident, and chairs. The participants performed the study in a sitting position. Figure 11 shows a schema of the experimental setup and pictures of the room where the experiment was performed.

Table 4 presents the step-by-step process of the experiment, which introduces the instruments used for inducing the independent variables and measuring the dependent variables. After the participants arrived in the experiment room, they received a short explanation about the experiment. They signed the informed consent form, and the eye-tracking system was calibrated. In order to get used to wearing the eye-tracking glasses, the participants completed the prequestionnaires while already wearing the eye-tracking glasses. After completing the prequestionnaires, the participants were given the instructions to the task and shown the picture of the fire. The instructions were as follows: “Please imagine being in the soccer stadium of Kaiserslautern. You see a fire on the opposite side of the stadium. You want to report this to the fire department using the mobile application. To give the fire department a better idea of the emergency, you decide to also send a picture of the fire. In the mobile application, you are going to see different screens. Please go through the application and try to use all features: forms, camera, chat function, location, etc.” After giving the instruction, the moderator handed a smartphone with the RESCUER App installed on it to the participant, who could then immediately start to report the fire with the RESCUER App.

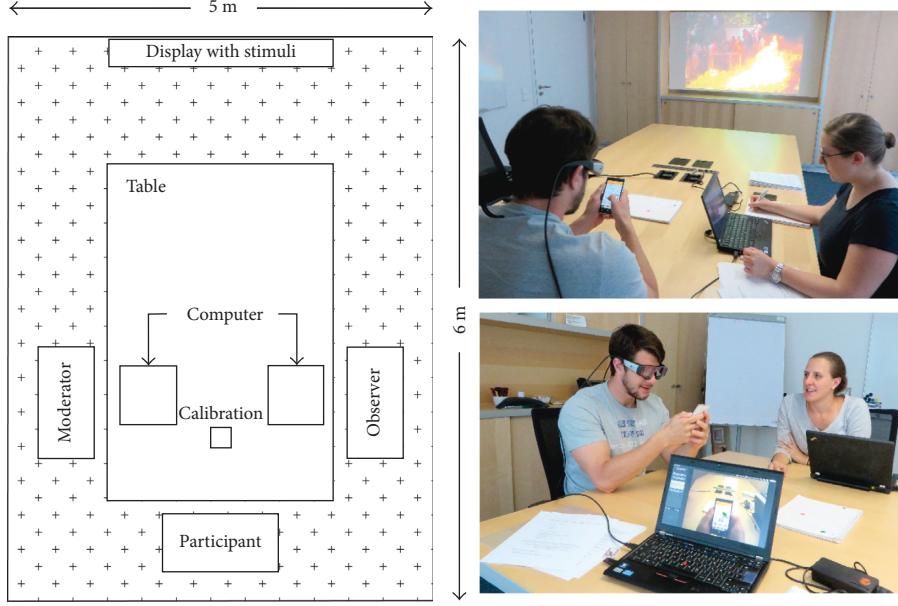


FIGURE 11: Experiment room and experimental setup.

TABLE 4: Experimental procedure.

Step	Stage	Activity
Step 1	Welcome	(i) Introduction to the experiment (ii) Participation consent (iii) Eye-tracker calibration
Step 2	Prequestionnaire	(i) Demographic information (ii) Experiences with previous emergencies
Step 3	Task 1	(i) Use of the RESCUER App (ii) Use of the eye tracker
Step 4	Post-task 1 questionnaire	(i) SMEQ for each screen (ii) SMEQ overall
Step 5	Task 2	(i) Use of the RESCUER App (ii) Use of the eye tracker (iii) Performance of the <i>n</i> -back task
Step 6	Post-task 2 questionnaire	(i) SMEQ for each screen (ii) SMEQ overall
Step 7	Postquestionnaire	(i) ISONORM 9241/10
Step 8	End	(i) Debriefing (ii) Explanation of study purpose

After performing task 1, the participants completed the SMEQ for each interaction mode and also provided an overall assessment of task 1 using the SMEQ. As soon as they had completed the questionnaires, they received follow-up instructions for task 2 and the audio-based one-back task was demonstrated. The instructions for task 2 were the same as for task 1, only with the addition of “Do not forget to perform the one-back task while reporting the emergency shown in the picture with the help of the RESCUER App.” After receiving the instructions, the participants performed task 2 describing a different picture of an emergency than the one they had described in task 1 (Figures 9 and 10). After completing this task using the RESCUER App, SMEQ questionnaires for each screen and an SMEQ questionnaire for the overall task were

once again completed by the participants. If the stress level was too high for the participants, they were allowed to interrupt the task. All participants were able to perform both tasks completely and without interruption.

Finally, the participants completed the ISONORM 9241/10 questionnaire and answered open-ended questions to provide feedback about their experience with RESCUER App. A detailed explanation about the study was offered to participants who wanted to know more about it.

6.6. Subjects. The participants of this study consisted of 19 male and 7 female employees from Fraunhofer IESE ($N=26$). The age of the participants ranged from 20 to 53 years

TABLE 5: Mean estimates of the SMEQ for all six experimental conditions ($n=26$).

Interaction mode (factor 1)	Level of cognitive load (factor 2)	Mean (M)	Standard error (SE)	95% confidence interval	
				Lower bound	Upper bound
One-click interaction	No/low load	19.50	3.18	12.96	26.04
	High load	44.23	6.46	30.92	57.54
Guided interaction	No/low load	69.15	8.81	51.00	87.30
	High load	122.19	9.73	102.15	142.23
Chat interaction	No/low load	105.58	10.94	83.04	128.11
	High load	166.27	7.51	150.80	181.74

($M = 32.77$, $SD = 7.63$). According to the Shapiro–Wilk test, age was distributed normally: $W(25) = 0.93$, $p = 0.09$. However, the distribution is skewed to the right, with skewness of 0.85 ($SE = 0.46$) and kurtosis of 1.27 ($SE = 0.89$). All the participants volunteered for the study and received no compensation.

The participants had a high level of education: 77% of the participants were graduates, 8% had a Ph.D., 11% had a degree in applied sciences, and 4% a secondary-level certificate. All the participants in the experiment possessed a smartphone: 53.84% used an iPhone, 34.62% were Android users, and a minority of 11.54% used Windows Phone.

With respect to prior experience with emergency situations, 42.31% of the participants had never been involved in an emergency situation, while 57.69% of the participants had already experienced an emergency situation (e.g., as an eye witness). The most typical emergencies were traffic accidents (40%), followed by fire (24%), medical emergencies (16%), and gas leakage (8%). Other kinds of incidents totaled 12% of the incidents. Of the 26 participants, 16 had never made an emergency call, six participants had made an emergency call twice, and three participants once. One of the participants was a member of a volunteer workforce and had performed an emergency call four times on duty.

7. Results

In this section, we report on the results separately for each of the research questions described in Section 6.1. First, we will analyze how the perceived mental effort when using the RESCUER App changes, taking into consideration the different interaction models and the levels of externally induced cognitive load (Section 7.1). Secondly, we will then check whether previous experience with emergency situations had an influence on the participants' experience of mental effort (Section 7.2). Thirdly, we will examine the duration of using the RESCUER App under different levels of cognitive load (Section 7.3). Fourthly, we will investigate the results of the one-back task, and how this objective measure relates to the subjective measure of perceived mental effort (Section 7.4). Lastly, we will examine the participants' overall usability evaluation of the RESCUER App based on items from the ISONORM 9241/10 questionnaire (Section 7.5). All statistical analyses were performed using SPSS 19. Unless otherwise stated, all statistical tests assume a significance level of $\alpha = 0.05$.

7.1. RQ1: Subjective Mental Effort of the RESCUER App under Different Levels of Cognitive Load. To analyze RQ1, we

performed a 2×3 repeated measurement ANOVA with two within-subject factors (interaction mode and cognitive load) for the dependent variable SMEQ. Table 5 shows the mean estimates for each combination of the levels of the two factors (interaction mode \times cognitive load).

Mauchly's test indicated that the assumptions of sphericity had been violated for the interaction effect of interaction mode (factor 1) and cognitive load (factor 2), $\chi^2(2) = 10.31$, $p = 0.006$, while for the main effects the assumption for sphericity was met (interaction mode: $\chi^2(2) = 0.65$, $p = 0.72$). Therefore, degrees of freedom were corrected for the interaction effect interaction mode \times cognitive load using Greenhouse–Geisser estimates for sphericity ($\varepsilon = 0.74$).

There was a significant main effect of the interaction mode on the rating of the SMEQ, $F(2, 50) = 71.14$, $p < 0.001$, $\eta_p^2 = 0.83$. Contrasts revealed that the ratings for screen 2 (guided interaction) were significantly higher than for screen 1 (one-click interaction), $F(1, 25) = 52.48$, $p < 0.001$, $r = 0.82$, and ratings for screen 3 (chat interaction) were significantly higher compared to screen 2, $F(1, 25) = 24.40$, $p < 0.001$, $r = 0.70$. There was also a significant main effect of the administered level of cognitive load on ratings of SMEQ, $F(1, 25) = 94.08$, $p < 0.001$, $\eta_p^2 = 0.79$. The contrast shows a significantly higher rating of SMEQ in the high cognitive load task compared to the low/no-load task, $F(1, 25) = 94.08$, $p < 0.001$, $r = 0.89$.

There was a significant interaction effect between the interaction mode and the level of cognitive load applied, $F(1.48, 37.06) = 7.11$, $p = 0.005$, $\eta_p^2 = 0.22$. This indicates that the interaction mode used by participants has different effects on participants' rating of SMEQ depending on the cognitive load a person experiences. Figure 12 shows an ordinal interaction between interaction mode (a) and level of cognitive load (b) for ratings of SMEQ. To break down this interaction, contrasts were performed that compared each interaction mode against its preceding mode with fewer interaction possibilities (one-click interaction versus guided interaction, and guided interaction versus chat interaction) for each level of cognitive load. The first contrast revealed a significant difference when comparing no/low cognitive load to high cognitive load for one-click interaction compared to guided interaction, $F(1, 25) = 18.76$, $p < 0.001$, $r = 0.65$. The remaining contrast shows no significant difference when comparing no/low cognitive load to high cognitive load for guided interaction versus chat interaction, $F(1, 25) = 0.42$, $p = 0.53$, $r = 0.13$. However, this contrast yields a small effect size.

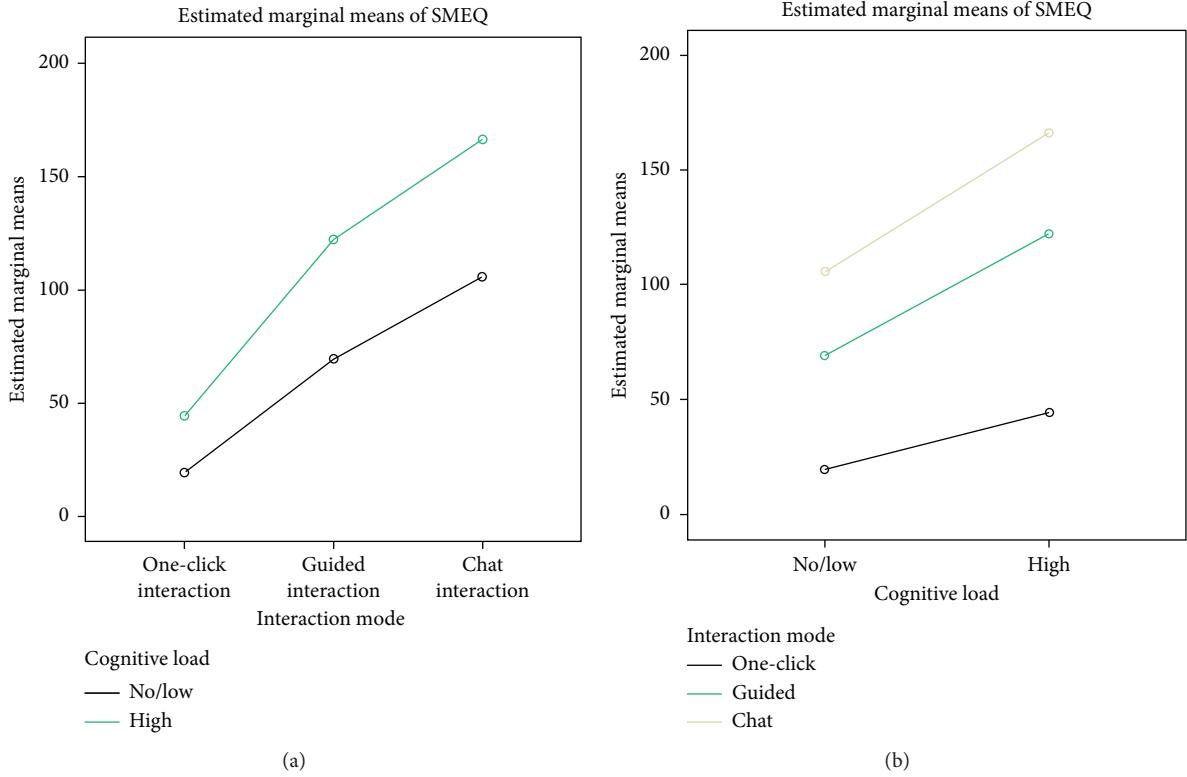


FIGURE 12: Interaction graphs for SMEQ. (a) Interaction mode. (b) Cognitive load.

TABLE 6: Mean estimates of SMEQ for all six experimental conditions separated according to previous experience with emergency situations.

Previous experience with an emergency (factor 3)	Interaction mode (factor 1)	Level of cognitive load (factor 2)	Mean (M)	Standard error (SE)	95% confidence interval	
			Lower bound	Upper bound		
No (<i>n</i> = 11)	One-click interaction	No/low load	18.36	4.51	8.32	28.41
		High load	32.64	4.86	21.80	43.47
	Guided interaction	No/low load	52.64	9.82	30.75	74.52
		High load	103.27	13.89	72.33	134.21
	Chat interaction	No/low load	85.91	14.21	54.26	117.56
		High load	170.55	8.51	72.33	134.21
Yes (<i>n</i> = 15)	One-click interaction	No/low load	20.33	4.53	10.63	30.04
		High load	52.73	10.23	30.80	74.67
	Guided interaction	No/low load	81.27	12.85	53.71	108.82
		High load	136.07	12.63	108.97	163.16
	Chat interaction	No/low load	120.00	15.16	87.49	152.51
		High load	163.13	11.62	138.22	188.05

7.2. RQ2: *Effect of Previous Emergency Experience on SMEQ while Using the RESCUER App under Different Levels of Cognitive Load.* In the analysis of this research question, we were interested in learning how previous experience with emergency situations might alter the effects of the interaction mode in SMEQ and experimentally induced cognitive load. Therefore, we included a third factor in the model, in which previous experience with emergency situations served as an additional independent variable (as a between-subject factor) in the ANOVA.

Of the 26 participants, 11 stated that they had no prior experience with an emergency situation, while 15 participants had already been involved in an emergency situation.

Table 6 shows the mean estimates for SMEQ for the six experimental conditions, separately for participants with previous emergency experience (lower part of the table) and without previous emergency experience (upper part of the table). Interestingly, participants with previous emergency experience rated SMEQ in all experimental conditions higher than participants without previous emergency experience, with one exception: for the chat interaction under high cognitive load, participants without emergency experience rated SMEQ higher.

Mauchly's test indicated that the assumptions of sphericity had been violated for the interaction effect of interaction mode (factor 1) and cognitive load (factor 2),

TABLE 7: Test results of the within-subject effects.

Source	F	df	df _{error}	p	η_p^2
Interaction mode	68.19	2	48	<0.001	0.74
Interaction mode × emergency experience	0.72	2	48	0.49	0.03
Cognitive load	91.69	1	24	<0.001	0.79
Cognitive load × emergency experience	0.43	1	24	0.52	0.02
Interaction mode × cognitive load ¹	10.03	1.52	36.41	0.001	0.30
Interaction mode × cognitive load × emergency experience ¹	5.56	1.52	36.41	0.01	0.19

¹Corrected degrees of freedom using Greenhouse–Geisser estimates for sphericity.

$\chi^2(2) = 8.82$, $p = 0.012$. Thus, degrees of freedom were corrected for the interaction effect interaction mode × cognitive load and interaction mode × cognitive load × emergency experience using Greenhouse–Geisser estimates for sphericity ($\epsilon = 0.76$).

The between-subject effect for emergency experience showed no significant result, $F(1, 24) = 3.36$, $p = 0.08$, $\eta_p^2 = 0.12$, indicating that participants with previous emergency experience and participants without previous emergency experience gave generally the same SMEQ ratings. The results for the within-subject and interaction effects are listed in Table 7.

The results show a significant main effect of interaction mode. Similarly to the previous analysis, contrasts show a significantly higher rating of the SMEQ for guided interaction compared to one-click interaction, $F(1, 23) = 49.28$, $p < 0.001$, $r = 0.83$, and a higher rating for chat interaction compared to guided interaction, $F(1, 23) = 25.55$, $p < 0.001$, $r = 0.73$. In addition, there is a significant main effect of cognitive load. Again the contrast shows that high cognitive load conditions yield a higher rating of SMEQ compared to no/low cognitive load conditions, $F(1, 24) = 91.69$, $p < 0.001$, $r = 0.89$.

There was no significant interaction effect for both interaction mode × emergency experience and cognitive load × emergency experience. However, the interaction mode × cognitive load interaction effect was significant, $F(1.52, 36.41) = 10.03$, $p = 0.001$, $\eta_p^2 = 0.30$. Correspondingly, this result indicates that the participants' SMEQ ratings differed for the three interaction modes depending on a person's cognitive load. To break down this interaction, contrasts were performed that compared each interaction mode against its preceding mode with fewer interaction possibilities (one-click interaction versus guided interaction and guided interaction versus chat interaction) for each level of cognitive load. The first contrast revealed a significant difference when comparing no/low cognitive load to high cognitive load for one-click interaction compared to guided interaction, $F(1, 24) = 19.83$, $p < 0.001$, $r = 0.67$. The second contrast shows no significant difference when comparing no/low cognitive load to high cognitive load for chat interaction when comparing it to guided interaction, $F(1, 24) < 1$, $p = 0.34$, $r = 0.19$. However, this contrast yields a small effect size.

Finally, the interaction mode × cognitive load × emergency experience interaction was significant, $F(1.52, 36.41) = 5.56$,

$p = 0.01$, $\eta_p^2 = 0.19$. This indicates that the interaction mode × cognitive load interaction described previously was different for participants without previous emergency experience and for participants with previous emergency experience. As can be seen in Figure 13, participants with previous emergency experience yielded a higher SMEQ rating in the chat interaction mode under high cognitive load, while for all other combinations of interaction mode and cognitive load, the SMEQ rating is lower compared to participants without previous experience. However, analysis of the contrasts shows that this effect might be overestimated because no contrast shows a significant result. The first contrast shows no significant difference between the one-click interaction mode and the guided interaction mode between the different levels of cognitive load for the two experience groups, $F(1, 24) = 1.12$, $p = 0.30$. The second contrast analyzes the difference between the guided interaction mode and the chat interaction mode between the different levels of cognitive load for participants with previous emergency experience versus participants without previous emergency experience, $F(1, 24) = 4.04$, $p = 0.06$. This contrast shows no significant result.

7.3. RQ3: Effect of Interaction Mode and Cognitive Load on Time Spent on Each Screen of the RESCUER App. In this research question, we were interested in investigating the participants' usage duration of each screen in seconds in order to describe the emergencies depicted in the stimulus pictures (Figures 9 and 10). To examine RQ3, we performed a repeated measures ANOVA with two within-subject factors (interaction mode and cognitive load). Due to technical problems with the eye tracker, three participants were eliminated from this analysis, resulting in $n = 23$.

Mauchly's test showed that the assumptions of sphericity had been violated only for the interaction effect of interaction mode (factor 1) and cognitive load (factor 2), $\chi^2(2) = 28.78$, $p < 0.001$. Consequently, degrees of freedom were corrected for the interaction effect interaction mode × cognitive load using Greenhouse–Geisser estimates for sphericity ($\epsilon = 0.57$).

There was a significant main effect of interaction mode on usage duration, $F(2, 44) = 216.73$, $p < 0.001$, $\eta_p^2 = 0.91$. Contrasts revealed that usage duration for screen 2 (guided interaction) was significantly longer than for screen 1 (one-click interaction), $F(1, 22) = 71.83$, $p < 0.001$, $r = 0.87$, and usage duration for screen 3 (chat interaction) was significantly

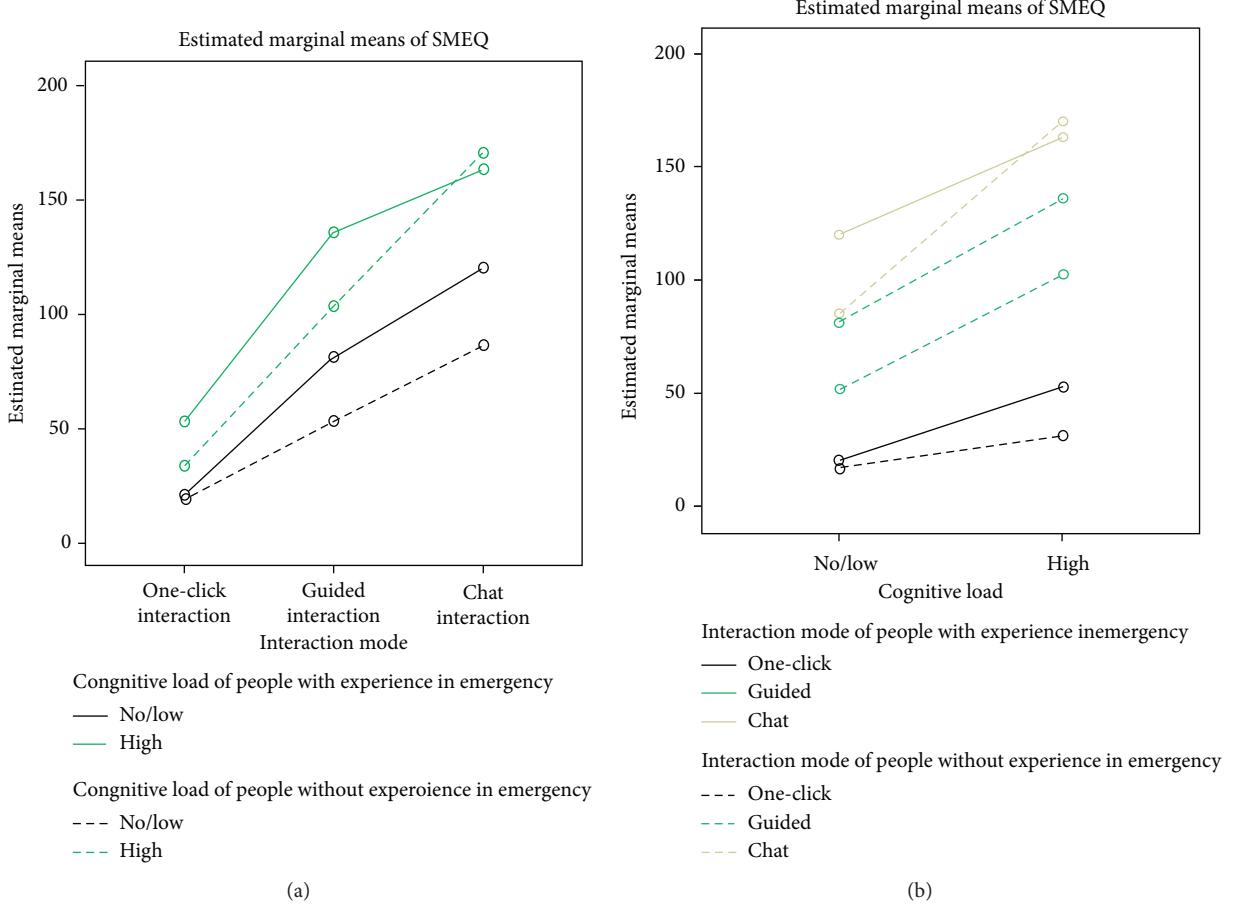


FIGURE 13: Interaction graphs for SMEQ for participants without previous emergency experience (dashed lines) and with previous emergency experience (continuous lines).

longer compared to screen 2, $F(1, 22) = 170.28$, $p < 0.001$, $r = 0.94$. There was also a significant main effect of the level of cognitive load applied on usage duration, $F(1, 22) = 54.38$, $p < 0.001$, $\eta_p^2 = 0.71$. Surprisingly, the contrast shows a significantly lower usage duration in the high cognitive load task compared to the low/no cognitive load task, $F(1, 22) = 54.38$, $p < 0.001$, $r = 0.84$.

Furthermore, there was a significant interaction effect between the interaction mode and the level of cognitive load applied, $F(1.15, 25.20) = 37.38$, $p < 0.001$, $\eta_p^2 = 0.63$. This indicates that the interaction mode used has different effects on the usage duration depending on a person's cognitive load. Figure 14 shows the interaction graphs for usage duration. To break down this interaction, contrasts were performed comparing each interaction mode against its preceding mode with fewer interaction possibilities (one-click interaction versus guided interaction and guided interaction versus chat interaction) separately for both levels of cognitive load. Only the contrast guided interaction versus chat interaction yielded a significant result, $F(1, 22) = 35.11$, $p < 0.001$, $r = 0.78$.

The results reflect that the interaction duration increases along the interaction modes, that is, screens with a larger number of functionalities logically have a longer usage duration. However, in contrast to our assumption, we also

see that usage duration significantly drops from task 1 (low/no cognitive load) to task 2 (high cognitive load) with the highest change in the chat interaction mode, possibly due to practice effects with the RESCUER App in the preceding task.

7.4. RQ4: Results of the n-Back Task. In the task 2 of the experiment, the participants had to perform an auditory n -back task conceived as a one-back task in parallel to using the mobile application. From the video recordings, we were able to extract the following variables to characterize the participants' performance with the parallel task:

- (i) The number of one-back stimuli a participant received in task 2, which depended on the interaction duration
- (ii) The number of correct responses to the one-back task (answering with "yes" when a stimulus was similar to the previous one or answering with "no" if two consecutive stimuli differed)
- (iii) The number of incorrect responses
- (iv) The number of omissions (when a participant skipped an answer).

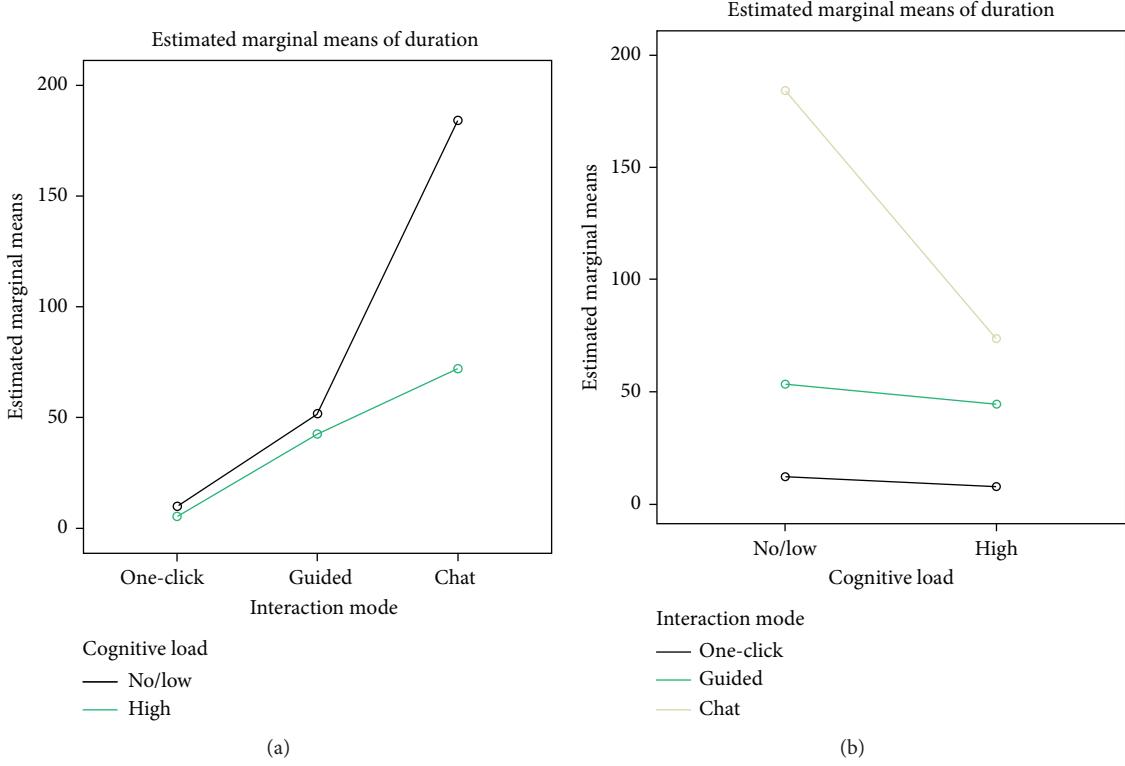


FIGURE 14: Interaction graphs for duration (in seconds).

Because the overall usage duration of the RESCUER App differed between participants, we calculated the ratio of the number of responses (correct, wrong, and omitted) and the number of the presented n -back stimuli for each participant. With this relative score, it is possible to perform a comparison across the study participants.

Overall, the participants provided an average of 86.61% correct responses ($SD = 13.67\%$) with a minimum score of 50% and a maximum score of 100% correct responses to the n -back task. This number significantly differs from the assumed 100% correct responses, $t(25) = -4.99$, $p < 0.001$. The remaining 13.39% are divided into incorrect responses (on average, 3.50% of all stimulus responses were incorrect, $SD = 4.47\%$) and omitted responses (on average, 9.89% of all stimulus responses were omitted responses, $SD = 13.37\%$). Therefore, a larger proportion of the noncorrect responses can be attributed to omitted responses, whereas incorrect responses have the smallest proportion. Still, both categories of noncorrect stimulus responses differ significantly from zero (incorrect responses: $t(25) = 3.99$, $p < 0.001$ and omitted responses: $t(25) = 3.77$, $p < 0.001$).

These results show that a large proportion of the given stimulus responses were correct answers, but a recognizable part of the stimulus responses were not correct, either because a wrong response was given or because no response was given at all because a participant missed a stimulus, possibly due to cognitive overload.

7.5. Usability of the RESCUER App. To analyze the participants' view on the usability of the mobile application, we used nine items from the ISONORM 9241/10 questionnaire

(Section 6.4). Table 8 lists the descriptive results, the one-sample t -test results (test value = 4; i.e., middle of response scale), and the frequency distribution for the answer categories for each item.

The participants rated items regarding learnability of the mobile application favorably. Ratings for item 7 ("The app requires just a little time to learn how to use it"), $t(25) = 15.01$, $p < 0.001$, item 8 ("The app is easy to learn without further assistance or a user manual"), $t(25) = 14.88$, $p < 0.001$, and the reversed item 9 ("The app requires that you have to remember many details"), $t(25) = -6.06$, $p < 0.001$, significantly differ from the middle of the response scale (the "neutral middle").

The participants strongly disagreed with the statement that the app "is complicated to operate," $t(25) = -4.57$, $p < 0.001$. Regarding the self-descriptiveness of the RESCUER App, on the one hand, the participants significantly consented that the app uses "understandable terms, abbreviations, or symbols in the screens and menus" (item 2), $t(25) = 7.99$, $p < 0.001$, but on the other hand, they disagreed in their opinion as to whether the app "provides information in adequate extent which entries are permitted or required" (item 3), $t(25) = 1.56$, $p = 0.50$. This apparent contradiction is in line with items regarding conformity with user expectations of the app, which include the ratings of items 4, 5, and 6. All three items measure the user's orientation in the app and feedback of the app to the users' input and revealed no significant positive or negative ratings. We assert that the lack of confirmation screens due to technical problems may have contributed to this dispersed rating.

Overall, the subjective ratings of items from the ISO-NORM questionnaire show a positive to neutral opinion on

TABLE 8: Results of the item selection from the ISONORM 9241/10 questionnaire.

Items	Descriptives						One-sample <i>t</i> -test (two-tailed; test value = 4)				Distribution
	<i>N</i>	<i>M</i>	SD	<i>T</i>	<i>df</i>	<i>p</i>	Mean difference	95% confidence interval of the difference			
								Lower	Upper		
(1) The app is complicated to operate.	26	2.85	1.29	-4.57	25	0.000	-1.154	-1.67	-0.63		
(2) The app uses easily understandable terms, names, abbreviations, or symbols in the screens and menus.	26	5.65	1.06	7.99	25	0.000	1.654	1.23	2.08		
(3) The app provides information in adequate extent which entries are permitted or required.	26	4.50	1.63	1.56	25	0.131	0.500	-0.16	1.16		
(4) The app facilitates orientation through a consistent look.	26	4.73	1.51	2.47	25	0.21	0.731	0.12	1.34		
(5) The app lets one in the dark about whether an entry was successful or not.	25	4.04	1.88	0.11	24	0.916	0.040	-0.74	0.82		
(6) The app sufficiently informs about what it is doing.	26	3.58	1.47	-1.46	25	0.156	-0.423	-1.02	0.17		
(7) The app requires just a little time to learn how to use it.	26	6.15	0.73	15.01	25	0.000	2.154	1.86	2.45		
(8) The app requires that you have to remember many details.	26	2.38	1.36	-6.06	25	0.000	-1.615	-2.16	-1.07		
(9) The app is easy to learn without further assistance or a user manual.	26	6.23	0.77	14.88	25	0.000	2.231	1.92	2.54		

M, mean; SD, standard deviation; ---, Does not apply at all; --, does not apply; -, rather does not apply; -/+ , undecided; +, rather applies; ++, applies; +++, applies fully.

the usability of the RESCUER App, and no explicit strongly negative usability assessment was observed.

8. Discussion

After having performed the user experiment and the statistical analyses, we obtained several insights regarding the interaction modes model as well as their underlying theoretical background and findings on how to improve the usability of the RESCUER App in later iterations. In this chapter, we will discuss the results presented above and the threats to validity that may have an influence on the main outcomes.

8.1. RQ1: Subjective Mental Effort of the RESCUER App under Different Levels of Cognitive Load. The analysis of the data for the first research question revealed that the participants reported higher mental effort (i.e., more difficulty) in operating the app under cognitive load than when they were not under cognitive load. In both conditions, the guided interaction (screen 2) was perceived as being much more difficult than the one-click interaction (screen 1) with the perceived difficulty from the guided interaction to the chat interaction (screen 3). This is surprising, as both screens 1 and 2 ask for simple input by clicking on buttons with predefined answers, whereas screen 3 requires formulating an answer to a question in natural language and could be hypothesized to be much more difficult. On the other hand, it is understandable that answering multiple

questions in one interface is more difficult than answering one open question, but answering multiple questions by clicking is still perceived as being easier than answering open questions through a chat interface, as it requires even more effort by the cognitive and motor processors [24].

This is also supported by two additional findings. First, the differences between the reported difficulty of the screens were smaller for the high cognitive load conditions, revealing that the mental effort required to perform simple tasks is impaired more than that for more difficult tasks. Second, the interaction between cognitive load and the difference between screens 2 and 3 did not reach significance, suggesting that performance is impaired equally for the more complex tasks, regardless of their form (e.g., answering multiple questions by clicking versus answering an open question through chatting). This is likely because more complex tasks already require more attention even under less cognitive load, so that other factors demanding a person's attention interfere but do not add to the complexity of the task itself. This suggests that people are generally capable of performing more complex tasks under stress without being impaired too much in executing these tasks. In other words, it is possible to use the RESCUER App in an emergency situation.

8.2. RQ2: Effect of Previous Emergency Experience on SMEQ while Using the RESCUER App under Different Levels of Cognitive Load. Comparing the SMEQ results of participants with or

without prior experience in emergency situations revealed that participants with such an experience showed a similar reporting pattern as participants without such an experience, but usually rated the screens more difficult in both types of mental load conditions. Although this difference did not reach significance, it did reveal a trend towards this difference. In addition, a significant interaction mode \times cognitive load \times emergency experience interaction showed that there is a difference in behavior depending on experience with emergencies, but an absence of other significant interactions with emergency experience reveals that this is based on a more complex interplay of factors. The notion that this might be caused by different expectancy of persons with experience regarding how the process is shaped can be refuted by the finding that the differences held true even in the second condition. Instead, the participants may be primed by their previous experience, and perhaps hindered in their performance by remembering a previous emergency situation they were in. According to the affect infusion model by Bower andForgas [37], affective states influence cognitive processes such as judgments. Therefore, it is possible that the question about previous emergency experience in the prequestionnaire primed an affective reaction, which in turn activated associations, evaluations, and judgments related to past emergencies that influenced the SMEQ rating in the experiment. However, these effects are small and did not dramatically reduce a participant's performance. This shows two things: people without prior experience with an emergency situation are no less able to successfully report this incident than people who have such an experience, and these people may in fact benefit from not being hindered by such an experience, even though this hindrance is fairly small. Furthermore, it shows that the RESCUER App can be used by anyone, regardless of prior experience.

8.3. RQ3: Effect of Interaction Mode and Cognitive Load on Time Spent on Each Screen of the RESCUER App. Similar to the results on reported effort (RQ1), the time spent per screen increases from screen 1 (one-click interaction) to screen 2 (guided interaction), and from screen 2 to screen 3 (chat interaction). On the SMEQ scale, the effort was perceived as higher under high cognitive load. The performance results in terms of task duration show a different pattern. The participants were quicker on each screen, especially on the third screen, for which they required a lot more time in the first condition. This similarity also becomes obvious from the difference between reported effort and screen duration. On the SMEQ scale, screen 2 was perceived as requiring much more effort than screen 1, more than the difference between screen 2 and screen 3. The results of the actual duration show that under no/low cognitive load conditions, the time spent on screen 3 was considerably higher than that of screen 2, but under high cognitive load conditions, the pattern of screen duration resembled that of reported effort, while the time spent on screens 1 and 2 slightly decreased and showed no significant interaction in terms of spending less time on either screen 1 or screen 2. These results can be explained by the strong effect of practice

in task 1 on the performance in task 2, from which it became clear that despite the impairment of additional effort due to a parallel task to increase cognitive load, the performance by participants was still better than in the first condition, where the participants were also still exploring and learning how to work with the app. This suggests that it would be advisable to encourage users to familiarize themselves with the app outside of an emergency context in order to enable them to perform well even when they are under stress in an actual emergency situation.

The results also show that the participants were surprisingly capable of typing under high cognitive load in the chat interaction (screen 3), experiencing an apparently low interference of the *n*-back task, where the participants had to indicate whether two subsequent tones were different or similar by verbal reporting (saying "yes" or "no," resp.). This means that under stress, multimodal processing is still possible and provides further evidence for the ability to operate the RESCUER App under high cognitive load.

8.4. RQ4: Results of the n-Back Task. Under high cognitive load conditions, the participants were instructed to focus on operating the RESCUER App while making as few errors in the parallel one-back task as possible. The results showed that the participants performed well. They navigated swiftly through the app, while still correctly responding to on average 86.61% of the stimuli. The participants who were able to respond correctly to 100% of the stimuli were musically trained and were able to both operate the app and perform the one-back task without much effort. The results of the participants' performance on the one-back task show that people are still able to process multiple sensory inputs (e.g., photo, app interface, and auditory stimuli) and outputs (e.g., tapping and typing) without too much impairment. Some information will be processed slower and some small lapses may occur, but the participants showed a remarkable ability to recuperate and resume both operating the app and attending to the stimuli.

8.5. Interaction Modes' Model and Theoretical Background. The results show that the effort experienced under high cognitive load can be significantly reduced if the number of questions is reduced and users can select pre-formatted answers. Persons under lower cognitive loads may be able to answer more questions, open questions, and more complex questions, but those directly involved in an emergency should be provided with simpler and pre-formatted questions. This could be achieved by distributing the questions that responders would like to ask over different persons within the crowd, for example, based on their distance from the site of the emergency or even on sensors in the smartphone. For example, movement sensors could help to identify whether a person is currently fleeing or shaking due to stress, in which case it would be better to provide clear instructions to the person on what to do rather than asking the person to answer a few questions. Furthermore, environmental factors and already collected answers could be used to limit the questions to those that still require answers,

thus further reducing the effort required from persons near an emergency site. People who have prior experience with emergencies do not have an advantage in using the RESCUER App. According to Breiner [25], a difference in use could be expected only if people are highly trained in using the RESCUER App so that they have the interaction flow stored in their long-term memory.

8.6. Usability Issues and Improvements. The outcomes of the ISONORM 9241/10 questionnaire are corroborated by the participants' verbal reports. Most of the participants indicated that they were missing feedback from the prototype of the RESCUER App during their interaction. The interactive prototype did not show a confirmation screen to the participants after each screen, such as those presented in Figure 8, screens (c) and (e). This means that the participants did not know if their report was sent. The final implementation of the RESCUER App includes such feedback to inform the users that their messages have been sent.

Another problem identified through the usability measurements was that sometimes the participants wanted to enter some information that was not covered by the multiple-choice answers. They were not sure if answering the question was mandatory or optional, and in that case, they mostly provided their best possible answer. In this respect, the RESCUER App could be improved by clarifying that there is no obligation to answer all questions, for example, by including evasive answer possibilities such as "I do not know," "I cannot say," "I am not sure," or "Pass," so that the user is freed from answering all questions presented to them.

The metaphor used for describing the magnitude of the fire (small, medium, and big) was often criticized by the participants because it left too much room for interpretation. Our recommended way of confirming whether the metaphor corresponds to reality is to perform a user study that compares a set of metaphors with the standard classification and checks whether users perceive them in the same way.

Finally, some users were bothered by some interaction concepts and the reaction of the RESCUER App's prototype if they did not conform to mobile platform interaction standards. The RESCUER App was designed to be a multi-platform mobile app, and the prototype was implemented as a web app. However, users of operating systems such as iOS or Android should be able to expect the RESCUER App to always behave in a similar way as other apps on their smartphone.

8.7. Threats to Validity. Although the study was prepared carefully, not all variables could be controlled for. Some aspects that might influence the degree to which the results and conclusions of this study can be confirmed with certainty and extrapolated to settings outside of this experiment were only recognized afterwards. In this section, we present the main threats to validity of our results.

8.7.1. Statistical Conclusion Validity. Before we conducted the study, we performed a power analysis with G*Power [38]

to calculate the optimal sample size so that effects could be detected. For each statistical analysis, the associated assumptions were tested (e.g., sphericity in repeated measures ANOVA) and, where necessary, the appropriate correction (e.g., Greenhouse-Geisser estimates) was used. However, the sample in our study was rather heterogeneous (e.g., wide age range) and therefore this might have had a negative influence on the detection of statistical effects and the interpretation of the results.

8.7.2. Internal Validity. The primary threat to the study's internal validity is certainly the effect of testing between the conditions no/low cognitive load and high cognitive load. Working with the app under the first condition meant that the participants had already been trained in the use of the app when they used it under the second condition. In this experiment, these conditions were not balanced across the participants, as the participants would likely have been hindered in their ability to learn how to use the app when under high cognitive load, and following this up with the low/no cognitive load condition would not have made any sense. It was of more interest to test whether the impairment under high cognitive load causes a person to work slower with the app than when first working with the app.

The use of the eye-tracking system, although non-obtrusive, somewhat narrows the visual field, which may have affected the participants' judgment in a negative sense [39]. Participants with glasses required greater adaptation to the eye-tracking system than users without glasses, which may have caused them to see the task as more difficult because they additionally had to strain their eyes. Although the eye-tracking system occasionally required recalibration, the analysis of the data included controlling for deviations in following the eye movements, meaning that the effect of instrumentation on validity was low.

The prototype used in the experiment was not fully functional and sometimes did not behave like a fully implemented mobile app. For example, clicking on text fields in order to input text required much more precision, and the map image on which participants could indicate their current position did not always appear. This behavior might have influenced the users' judgment regarding the usability of the RESCUER App, and the unexpected app behavior may have increased the participants' stress level during the use of the application.

8.7.3. Construct Validity. Stress was induced and cognitive capability reduced through the *n*-back task, causing multiple sensorimotor channels to require resources from working memory for the execution of the primary task in combination with the *n*-back task. Although this was an intended effect, it is an abstraction from real-world settings. When people are confronted by an actual threat, the load on their working memory may be different from that of the *n*-back task, as the stimuli will not always be auditory and hence may lead to different mental ability.

The presence of two experimenters may have caused an experimenter effect, meaning the participants may have

behaved differently than if they had used the app by themselves. In addition, as both the experimenters and the participants are employees of the same institute, they often knew each other well. Although the participants were motivated to help improve the app and the institute has an open and honest culture in which criticism may be expressed, the familiarity may have given rise to participants filling out the questionnaire in a more positive or socially desirable way. We believe this effect to be stronger than the possible negative mood induced by the use of an eye-tracker system, as the participants could nevertheless consciously choose to answer questions in a certain way, regardless of their actual mood.

Although we expected the requirements specified for the interaction modes to be appropriate for an emergency in which the difficulty of an interface increases gradually, adapting to the current state of the individual, it was uncertain whether the sequence from guided interaction to chat interaction was appropriate, especially considering the widespread use of mobile apps for chatting. As a consequence, the participants were highly trained in this kind of interaction so that the chat interaction could perhaps be perceived as being easier than the guided interaction. However, based on the results, we can assume that the content provided by the users during the chat interaction also influenced their cognitive capability. Nevertheless, putting the screens in the same order could have influenced the judgment of the participants regarding the difficulty of each screen and consequently may have confounded the results of the user experiment.

8.7.4. External Validity. As all participants were employees of the institute where the system was conceived, and as most had a background in computer science, there was a selection bias. Most participants are used to working with prototypes and thinking critically about interaction design and do not shun technology. This could have interacted with the experimental variable of the time required to understand the application and solve (intended and unintended) problems that occurred. None of the participants were directly involved in the development, but a primary identification with the institute can also influence how people assess the RESCUER App. The number of men participating was much higher than the number of women, which could have biased the processing of the interface in favor of the way men process information. A more balanced sample with respect to educational level might, among other things, have caused the effects of cognitive load during the guided interaction to be more severe, and the interpretation of metaphors and the description of the fire might have been different.

Moreover, this study did not evaluate the RESCUER App in a real emergency situation, but rather in a laboratory setting. Simulating a stress situation by increasing the cognitive load of the users is far different from the stress caused by a real emergency. A lot of factors related to an emergency situation were not replicated in this study, such as the crowd of people, the fact that people are moving while interacting with the app, and the threat itself. As a result, the

stress that was induced may be limited to only working under high cognitive load (or occasionally cognitive overload) and might not have led to a fight-or-flight reaction, as could be expected to occur during an actual emergency situation. To increase the external validity of our user experiment, a more realistic emergency setting would be required, although the development of an experiment in which life and property could be endangered in order to prove our hypotheses would be subject to many ethical concerns.

9. Conclusion and Future Work

This paper has described a human reaction model for the event of an emergency. Three interaction modes were derived from this model: one-click interaction, guided interaction, and chat interaction. The requirements defined for these interaction modes were used to develop a prototype of a mobile application, called the RESCUER App. The RESCUER App supports a command center in gathering information about an emergency with the help of the crowd that is involved in the emergency, as well as in giving instructions to the crowd during the emergency. The RESCUER App is a crowdsourcing-based emergency mobile application that is especially designed for emergency situations in which a large number of people are involved, such as at a festival or in a large industrial park. In particular, the RESCUER App was designed to allow reporting of incidents at the incident site when people are still under stress.

An experiment was performed to verify whether the interaction modes and the concrete mobile app design support the interaction of people with the RESCUER App when they are placed under stress. In the experimental design, the participants first used an interactive prototype of the RESCUER App with no/low cognitive load and a second time under high cognitive load. This high cognitive load was induced by having the participants perform an n -back task in addition to using the app.

The results show that people are remarkably capable of operating the app under high cognitive load, and that this ability does not depend on prior experience with emergency situations. For the three modes of interaction, the results showed that the one-click interaction mode demands very low cognitive load from people, the guided interaction mode generates moderate cognitive load by people, and the chat interaction causes a high mental effort for people. This indicates that the one-click interaction mode and an improved version of the guided interaction mode may be good ways to gather information from the crowd in situations where they experience higher levels of stress. Whenever users have a lower level of stress and require fewer resources for dealing with the emergency, the chat interaction mode can also be used and would allow sharing more detailed information such as the potential cause of the incident and personal impressions. Performance in all of these modes, even under stress, is better when people have familiarized themselves with the reporting process in advance, so it would be helpful to provide a simulation mode in which a user can test what it is like to report an emergency using the RESCUER App.

Improvement of the guided interaction could be achieved by reducing the number of questions presented to the user and spreading subsets of questions over individuals in the crowd, either randomly or according to some metrics. In addition, the use of validated metaphors for the multiple-choice questions (e.g., unambiguous icons) could facilitate easy and quick use of the app and lead to a reduction of cognitive load in the guided interaction. The design of these multiple-choice questions should be validated through one or more studies to compose a question catalog for use by operational forces and presentation to the crowd during an emergency. Other ways to improve the interaction are to make it more self-evident that answering questions is not mandatory and to improve the app's overall usability.

This study has the potential for improving crowd safety in emergencies to the extent that people in the crowd using a mobile application as the one proposed in this paper can notify a command center of an emergency at the first opportunity (as soon as they are able to press a button) without exposing them to further risks (as they can go away from the emergency site and use the application according to their cognitive capabilities), and without any unnecessary delays (without the need for looking for security staff at the venue or for waiting their phone call to be answered). The command center in turn can react faster and better to the situation due the richer set of information available at an earlier point in time. This includes the possibility of promptly sending messages to the crowd to guide them in how to behave or where to go, which again has a positive impact on the crowd's safety in emergencies.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

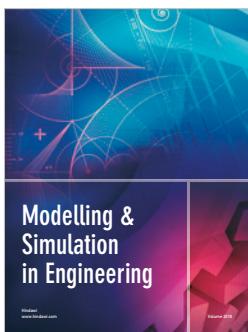
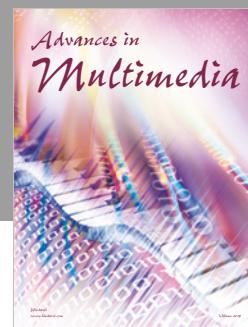
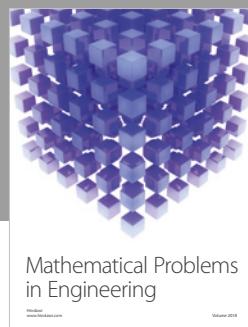
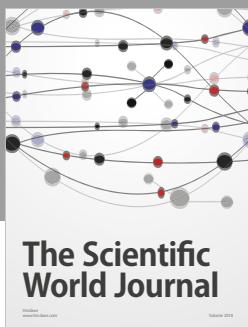
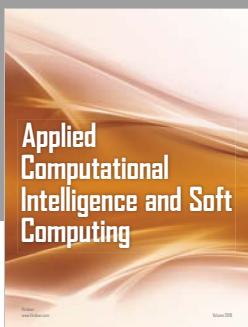
Acknowledgments

This work is supported by the RESCUER project, funded by the European Commission (Grant no. 614154) and by the Brazilian National Council for Scientific and Technological Development (CNPq/MCTI; Grant no. 490084/2013-3). The authors thank all participants in the experiment for their contributions in this project. Furthermore, the authors would like to thank Sonnhild Namingha for proofreading an earlier version of this paper.

References

- [1] W. Challenger, W. C. Clegg, and A. M. Robinson, *Understanding Crowd Behaviours: Guidance and Lessons Identified*, UK Cabinet Office, London, UK, 2009.
- [2] Association of Chief Police Officers (ACPO), "Guidance on emergency procedures," 2009.
- [3] D. A. McEntire, *Wiley Pathways Disaster Response and Recovery*, Wiley, Hoboken, NJ, USA, 2006.
- [4] E. L. Quarantelli, "Sociology of panic," in *International Encyclopedia of the Social and Behavioural Sciences*, N. J. Smelser and P. B. Baltes, Eds., pp. 11020–11023, Elsevier: Q5 Pergamon Press, New York, NY, USA, 2001.
- [5] J. Drury, C. Cocking, and S. Reicher, "Everyone for themselves? A comparative study of crowd solidarity among emergency survivors," *British Journal of Social Psychology*, vol. 48, no. 3, pp. 487–506, 2009.
- [6] S. M. Garcia, K. Weaver, G. B. Moskowitz, and J. M. Darley, "Crowded minds: the implicit bystander effect," *Journal of Personality and Social Psychology*, vol. 83, no. 4, pp. 843–853, 2002.
- [7] E. Riddell, A. Rock, P. Clothier, and J. London, *Strengthening Youth Resilience to Natural Disaster with Smartphone Technology: An Exploration of the Communication Needs, Priorities and Preferences of Queensland Youth for Improved Resilience to Natural Disasters and Recommendations for an Appropriate App*, p. 79, University of Queensland, School of Journalism and Communication and the Centre for Educational Innovation and Technology, Queensland, Australia, 2011.
- [8] N. Morrow, N. Mock, A. Papendieck, and N. Kocmich, "Independent evaluation of the Ushahidi Haiti project," *Development Information Systems International*, vol. 8, p. 2011, 2011.
- [9] Gordon College CCS, "Mitigating disaster risks using a crowd sourcing cross platform mobile apps," 2014, <https://www.climatecolab.org/contests/2014/crowdsourcing-for-disaster-risk-management/c/proposal/1306406>.
- [10] T. A. Kass-Hout and N. di Tada, "International system for total early disease detection (InSTEDD) platform," *Advances in Disease Surveillance*, vol. 5, no. 2, p. 108, 2008.
- [11] InSTEDD, "InSTEDD innovative support to emergencies," 2014, <http://instedd.org>.
- [12] K. N. Duc, T. T. Vu, and Y. Ban, "Ushahidi and Sahana Eden open-source platforms to assist disaster relief: geospatial components and capabilities," in *Geoinformation for Informed Decisions*, pp. 163–174, Springer International Publishing, Berlin, Germany, 2014.
- [13] K. Villela, K. Breiner, C. Nass, M. Mendonca, and V. Vieira, "A smart and reliable crowdsourcing solution for emergency and crisis management," in *Proceedings of 22nd Interdisciplinary Information Management Talks (IDIMT 2014)*, Networking Societies–Cooperation and Conflict, Podebrady, Czech Republic, September 2014.
- [14] X. Pan, "Computational modeling of human and social behaviors for emergency egress analysis," Ph.D. thesis, Stanford University, Stanford, CA, USA, 2006.
- [15] P. Gantt and R. Gantt, "Disaster psychology dispelling the myths of panic," *Professional Safety*, vol. 57, no. 8, pp. 42–49, 2012.
- [16] R. Perry and M. Greene, "The role of ethnicity in the emergency decision-making process," *Sociological Inquiry*, vol. 52, no. 4, pp. 306–334, 1982.
- [17] J. Leach, "Why people 'freeze' in an emergency: temporal and cognitive constraints on survival responses," *Aviation, Space and Environmental Medicine*, vol. 75, no. 6, pp. 539–542, 2004.
- [18] M. E. Kemeny, "The psychobiology of stress," *Current Directions in Psychological Science*, vol. 12, no. 4, pp. 124–129, 2003.
- [19] M. A. Staal, *Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework*, NASA Technical Memorandum 2004–212824, Washington, DC, USA, 2004.
- [20] M. M. Parker, *Understanding Psychology*, BVT Publishing, Redding, California, USA, 2010.

- [21] R. Plutchik, "Emotion: theory, research, and experience," in *Theories of Emotion 1*, Vol. 1, Academic Press, New York, NY, USA, 1980.
- [22] L. Minh, C. Adam, R. Canal, B. Gaudou, H. Tuong Vinh, and P. Taillandier, "Simulation of the emotion dynamics in a group of agents in an evacuation situation," in *Principles and Practice of Multi-Agent Systems*, N. Desai, A. Liu, and M. Winikoff, Eds., pp. 604–619, Springer, Berlin, Germany, 2012.
- [23] S. K. Card, T. P. Moran, and A. Newell, "The model human processor: an engineering model of human performance," in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufmann, and J. P. Thomas, Eds., p. 45, Wiley and Sons, New York, NY, USA, 1986.
- [24] W. B. Verwey, "Processing modes and parallel processors in producing familiar keying sequences," *Psychological Research*, vol. 67, no. 2, pp. 106–122, 2003.
- [25] K. Breiner, "AssistU—a framework for user interaction forensics," Ph.D. thesis, University of Kaiserslautern, Kaiserslautern, Germany, 2013.
- [26] F. Zijlstra, "Efficiency in work behavior. A design approach for modern tools," Ph.D. thesis, Delft University of Technology, Delft, Netherlands, 1993.
- [27] P. Beavon, "Improving memory through N-back training," Edith Cowan University, School of Psychology and Social Science Presentations, Queensland, Australia, 2012.
- [28] A. Baddeley, "Working memory and language: an overview," *Journal of Communication Disorders*, vol. 36, no. 3, pp. 189–208, 2003.
- [29] G. Schuhfried, *Wiener Testsystem*, Vienna Reaction Unit, Basic Program. Development and Production of Scientific Equipment, Mödling, Austria, 1992.
- [30] D. Mazzoni, M. Brubeck, and J. Haberman, "Audacity: free audio editor and recorder," 2005, <http://audacity.sourceforge.net>.
- [31] L. Turner, "A flare for soccer: open flames in soccer stadiums worldwide," 2009, http://www.boston.com/sports/blogs/bigshots/2009/05/a_flare_for_soccer_open_flames.html.
- [32] Wikimedia, "Wikimedia Commons the free media repository, Maccabi Haifa Supporters during the Qualifier vs. FC Salzburg," 2009, https://commons.wikimedia.org/wiki/File:Bengalische_Feuer,_Maccabi_Haifa_Fans_im_EM-Stadion_Wals-Siezenheim.jpg.
- [33] RESCUER App, "Interactive prototype," June 2017, <http://xudtxg.axshare.com/home.html>.
- [34] M. Schuette, "Bestimmung der bedingungsbezogenen Messgenauigkeit der Anstrengungsskala. Determination of the condition-related reliability of the Anstrengungsskala (Effort Scale)," *Zeitschrift fuer Arbeitswissenschaft*, vol. 56, no. 1-2, pp. 37–45, 2002.
- [35] W. Albert and T. Tullis, *Measuring the User Experience*, Morgan Kaufmann, Burlington, MA, USA, 2010.
- [36] J. Prümper, "Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität," in *Software-Ergonomie'97*, pp. 253–262, Vieweg+ Teubner Verlag, 1997.
- [37] G. H. Bower and J. P. Forgas, "Affect, memory, and social cognition," in *Cognition and Emotion*, E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas, and P. M. Niedenthal, Eds., pp. 87–168, Oxford University Press, New York, NY, USA, 2000.
- [38] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses," *Behavior Research Methods*, vol. 41, no. 4, pp. 1149–1160, 2009.
- [39] R. Verleur, "Affective video on the web: exploring presentation and task effects," Doctoral dissertation, University of Twente, Enschede, Netherlands, 2008.



Hindawi

Submit your manuscripts at
www.hindawi.com

