Integrating Affective Expressions into Robot-Assisted Search and Rescue to Improve Human-Robot Communication

by

Sami Alperen Akgun

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master's of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that this thesis consists of material all of which I authored or co-authored: see Statement of Contributions. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Conference Paper:

- Sami Alperen Akgun, Moojan Ghafurian, Mark Crowley, and Kerstin Dautenhahn. 2020. Using Emotions to Complement Multi-Modal Human-Robot Interaction in Urban Search and Rescue Scenarios. In Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20). Association for Computing Machinery, New York, NY, USA, 575–584. DOI:https://doi.org/10.1145/3382507.3418871 [3]

Workshop Paper:

- Sami Alperen Akgun, Moojan Ghafurian, Mark Crowley and Kerstin Dautenhahn. 2021. Integrating Affective Expressions into the Search and Rescue Context in order to Improve Non-Verbal Human-Robot Interaction. In Exploring Applications for Autonomous Non-Verbal Human-Robot Interactions Workshop of ACM/IEEE International Conference on Human-Robot Interaction 2021 (HRI'21) [4]

Sami Alperen Akgun, Prof. Moojan Ghafurian, Prof. Mark Crowley and Prof. Kerstin Dautenhahn designed the study (referred as Experiment 1 in this thesis) [3]. Sami Alperen Akgun conducted the study, the data analysis and wrote the first draft of manuscripts. Prof. Moojan Ghafurian, Prof. Mark Crowley and Prof. Kerstin Dautenhahn supervised the project. They reviewed and edited the manuscripts. Material from this study was used in Chapter 3.

Sami Alperen Akgun drafted the workshop paper that explains the general research approach taken to integrate emotions to Search and Rescue (SAR) context [4]. Moojan Ghafurian, Mark Crowley, and Kerstin Dautenhahn reviewed and edited the manuscript. Material from this short workshop paper was used in the introduction (Chapter 1) and background literature (Chapter 2).

Other Contributions:

Material from these contributions is not covered in this thesis since they are not related to the focus of this thesis.

- Moojan Ghafurian, Sami Alperen Akgun, Mark Crowley, and Kerstin Dautenhahn. 2021. Recognition of a Robot's Affective Expressions under Conditions with Limited Visibility. To appear in Proceedings INTERACT 2021, the 18th International Conference promoted by the IFIP Technical Committee 13 on Human–Computer Interaction, August 30 - September 3, Bari, Italy. Springer, Cham. [64]

  Sami Alperen Akgun, Prof. Moojan Ghafurian, Prof. Mark Crowley, and Prof. Kerstin Dautenhahn designed the study [64]. Prof. Moojan Ghafurian conducted the study and did the data analysis with Sami Alperen Akgun's assistance. Prof. Moojan Ghafurian wrote the majority of the first draft, while Sami Alperen Akgun wrote the remainder. Prof. Mark Crowley and Prof. Kerstin Dautenhahn reviewed and edited the manuscripts and supervised the project. Detailed of this study are not covered in this thesis, since this research was related to, but not central to this MASc research program.

- Hamza Mahdi, Sami Alperen Akgun, Shahed Saleh, and Kerstin Dautenhahn. A survey on design and evolution of social robots — past, present and future. Submitted to journal, 2021.

  Sami Alperen Akgun co-authored the survey paper which is currently under review.

- Austin Kothig, John Munoz, Sami Alperen Akgun, Alexander M. Aroyo, and Kerstin Dautenhahn. Connecting humans and robots using physiological signals – closing-the-loop in HRI. To appear in IEEE RO-MAN 2021, the 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, 2021 [110].

  Sami Alperen Akgun assisted Austin Kothig to implement software for Robot Operating System (ROS) in [110].

Abstract

Unexplained or ambiguous behaviours of rescue robots can lead to inefficient collaborations between humans and robots in robot-assisted SAR teams. To date, rescue robots do not have the ability to interact with humans on a social level, which is believed to be an essential ability that can improve the quality of interactions. This thesis research proposes to bring affective robot expressions into the SAR context to provide rescue robots social capabilities.

The first experiment presented in Chapter 3 investigates whether there is consensus in mapping emotions to messages/situations in Urban Search and Rescue (USAR) scenarios, where efficiency and effectiveness of interactions are crucial to success. We studied mappings between 10 specific messages, presented in two different communication styles, reflecting common situations that might happen during search and rescue missions and the emotions exhibited by robots in those situations. The data was obtained through a Mechanical Turk study with 78 participants. The findings support the feasibility of using emotions as an additional communication channel to improve multi-modal human-robot interaction for urban search and rescue robots and suggest that these mappings are robust, i.e., are not affected by the robot's communication style.

The second experiment was conducted on Amazon Mechanical Turk as well with 223 participants. We used Affect Control Theory (ACT) as a method for deriving the mappings between situations and emotions (similar to the ones in the first experiment) and as an alternative method to obtaining mappings that can be adjusted for different emotion sets (Chapter 4). The results suggested that there is consistency in the choice of emotions for a robot to show in different situations between the two methods used in the first and second experiment, indicating the feasibility of using emotions as an additional modality in SAR robots.

After validating the feasibility of bringing emotions to SAR context based on the findings from the first two experiments, we created affective expressions based on Evaluation, Potency and Activity (EPA) dimensions of ACT with the help of LED lights on a rescue robot called Husky. We evaluated the effect of emotions on rescue workers' situational awareness through an online Amazon Mechanical Turk Study with 151 participants (Chapter 5). Findings indicated that participants who saw Husky with affective expressions (conveyed through lights) had better perception accuracy of the situation happening in the disaster scene than participants who saw the videos of the Husky robot without any affective lights. In other words, Husky with affective lights improved participants' situational awareness.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Kerstin Dautenhahn for believing my abilities and giving me the honor to work with her. I am also grateful to my co-supervisor Prof. Moojan Ghafurian for constantly supporting me throughout this challenging journey. Without your support Prof. Kerstin Dautenhahn and Prof. Moojan Ghafurian, I would not be able to succeed. I hope our paths will cross again in the future.

I would like to thank Prof. Jennifer Boger for allocating time to read my thesis and providing valuable feedback.

I am grateful to share the same office with my pals Ellie Sanoubari and Austin Kothig. Thank you for filling pre-covid times with amazing memories. Ellie, I particularly appreciate your help regarding my research and always providing insightful comments. I would also like to thank Hamza Mahdi, Pourya Aliasghari and Shahed Saleh for their contributions in my work. It was a joy to collaborate with you.

Lastly, I would like to express my gratitude to my friends that I have known for years and the ones I met in Canada. Thank you for always being there for me whenever I needed it.

This research was undertaken, in part, thanks to funding from the Canada 150 Research Chairs Program.

Dedication

This thesis is dedicated to my family. Thank you, dad, mom, and brother. I would not be the person I am today if it was not for you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACT Affect Control Theory v, 3–5, 11, 12, 28, 31, 37, 38, 40, 43, 44, 48, 72, 75, 76

AIC Akaike's Information Criterion 54, 57

EPA Evaluation, Potency and Activity v, 11, 12, 28, 29, 31–38, 40, 72, 75, 76

HCI Human-Computer Interaction 3, 13, 16, 40, 76

HRI Human-Robot Interaction 3, 5, 9, 10, 13, 16, 40, 76

LMM Linear Mixed-effects Model 53, 54, 57, 58, 61, 67

PTSD Post Traumatic Stress Syndrome 37

SAR Search and Rescue iii, v, 1–13, 15, 16, 28–30, 32, 37–40, 51, 54, 55, 58, 59, 65, 67, 70–72, 74–77

USAR Urban Search and Rescue v, xi, 1, 6, 8, 9, 14–18, 23–27, 29, 74, 76

Chapter 1

Introduction

Emergencies that require search and rescue operations have been increasing in number every year [58]. These situations may happen due to natural or man-made [188] causes and need an immediate response, as time is a crucial element for the success of SAR operations [1]. Therefore, improving the efficiency of communication in SAR teams can be beneficial for the success of time-critical SAR operations.

In this thesis, we investigated the usage of affective expressions [1] in robot-assisted SAR. The following research questions were addressed through online studies where participants were presented with different SAR scenarios.

RQ1 Is there a consensus on what emotions to be used by USAR robots when they try to convey information about the situations commonly occurring during USAR operations? (Chapter 3)

RQ2 Is the mapping between emotions and USAR situations robust and not dependent on the wording of the sentences? (Chapter 3)

RQ3 How can a mapping between SAR-related sentences and emotions be obtained, and is there a way to generalize such mapping without limiting it to a specific set of emotions? (Chapter 4)

RQ4 How can affective expressions be designed and implemented on appearance-constrained SAR robots using lights? (Chapter 5)

[1] As defined in [87], affective expressions are the expressions that people use to show how they feel and what they believe. In this work, affective expressions are sometimes referred to as "emotions".

RQ5 Can affective expressions complement and improve multi-modal communication in human-robot SAR teams? (Chapter 5)

1.1 Motivation

Although rescue robots have been used in SAR operations since early the 2000s [28], they still need external help to operate properly. To the best of our knowledge, to date, there are no fully autonomous rescue robots that can work in unstructured and cluttered real-life SAR operations [49]. However, rescue robots can still act as teammates and improve human field workers' efficiency. To that end, a high level of collaboration between human-robot teammates should be achieved, which requires the implementation of clear and natural communication channels between the human and robot teammates. Unfortunately, human-robot interaction has been a bottleneck in robot-assisted SAR operations [28, 49]. In many situations, the intention behind robot teammates' actions is unclear to the field workers, i.e., they do not know what the robot is doing or why it is behaving in a specific way. This lack of transparency in robot teammates' behaviour has been identified as the main reason for inefficiency in SAR teams [111]. Therefore, using affective communication between human field works and rescue robots by taking advantage of multi-modal communication and developing alternative robot-to-human communication modalities would help overcome this bottleneck in robot-assisted SAR operations. This social ability of robots might also help victims in SAR situations who encounter robots to feel calmer until the medical treatment team arrives, preventing a shock [126, 20].

Most of the rescue robots used today are already equipped with different communication modalities such as voice, text, photos, and videos [92]. Nonetheless, these modalities may not be enough to provide efficient communication in human-robot SAR teams. For instance, voice is not effective for most of the SAR operations because rescue scenes are often noisy [120, 20]. Modalities other than voice can work in noisy environments, but they put the extra mental workload on field workers, or they do not work well depending on the search scene due to technical problems like delays and interferences [92]. Hence, combining different communication modalities can add redundancy and create robust communication in human-robot SAR teams to ensure that the others can be used as alternatives if one of the modalities stops working accurately. In this thesis, we propose using emotions and other affective expressions in a communicative way to complement existing communication modalities in human-robot SAR teams.

Since people are skilled at perceiving basic emotions without any training [43], and this process is intuitive, so does not require significant mental workload [205]. This makes using emotions a good modality to complement the existing multi-modal communication methods used in SAR robots. Employing this modality could contribute to overcoming the present problems in SAR robots related to interaction among teammates (humans and robots). It offers a way to reduce the cognitive load of human teammates to understand robot teammates' behaviour during SAR operations [106]. Providing a way for robots to express emotions will also give SAR robots an ability to interact socially with humans that would help SAR teams to operate in a more natural and efficient way [20]. Furthermore, this social ability is necessary to build affective robots in order to communicate with humans more naturally [184, 51, 96, 100].

1.2 Summary of Contributions

The following contributions are made through the research presented in this thesis:

1. The first experiment introduces affective expressions as a way to notify rescue workers during SAR missions. The findings of the first experiment support the feasibility of using emotions to convey information from rescue robots to workers.

2. The second experiment complements the findings of the first experiment. The process of matching SAR related scenarios with emotions was automated and less error-prone in this experiment. It was achieved by making use of EPA dimensions suggested by ACT.

3. The last experiment implements affective expressions on an appearance-constrained rescue robot (Clearpath Robotics Husky) to test the idea of using emotions to notify rescue workers. Results of the experiment suggested that participants who saw a rescue robot with an ability to express emotions had a better situational awareness.

1.3 Thesis Overview

Background Literature related to SAR, robot-assisted SAR, Human-Robot Interaction (HRI) in SAR, affective expressions in Human-Computer Interaction (HCI) and HRI, ACT and sentiment analysis was explained in Chapter 2. Chapter 3 explains the first experiment where we investigated the feasibility of using affective expressions for communication in robot-assisted SAR teams. The next study presented in Chapter 4 complements the findings of the first study, and it introduces ACT and EPA dimensions for communication in SAR context. Chapter 5 demonstrates the findings of the last experiment in this thesis, where affective expressions were designed using lights and implemented on a rescue robot. The effect of these emotional light displays on participants' situational awareness of the disaster scene was examined. Lastly, the conclusion was made in Chapter 6 as well as limitations of research presented in this thesis. Ethics clearance certificates and online interfaces used in the studies were given as an appendix (Appendix A, B, C and D).

Chapter 2

Background Literature

The ultimate goal of the research presented in this thesis is to complement the existing multi-modal interaction capabilities of rescue robots (voice, text, images, and videos [92]) with the use of affective expressions as an additional communication channel. The focus is on notifying rescue workers about common situations that might happen during SAR missions by using affective expressions to convey information from robots to humans better. In this way, rescue robots will be able to offer an intuitive (since emotions are instinctive to understand) and robust (due to redundancy in communication channels) way to notify field workers who might be overwhelmed by the stress of the rescue environment and complex inputs [92]. Such a communication system can help improving interactions between human and robot teammates in SAR. To the best of our knowledge, this is the first time that robot emotions are proposed as a complementary modality in SAR applications to improve the communication between human and robot teammates. In this background section, we will first introduce SAR and then describe the existing work on using robots in SAR situations, HRI in existing SAR robots, affective expressions in HRI, and ACT as well as some work on sentiment analysis that is related to this work.

2.1 Search and Rescue (SAR)

SAR is the general term for searching for people who are lost, trapped, and (might be) in danger. It is a broad term and has many sub-fields, usually depending on the area that rescue workers are looking for, which includes the following situations:

- **Mountain Rescue**: It usually occurs in mountains and differs from the other types of SAR operations in the sense that most of the mountain rescue victims are injured, ill, or lost [95], and they require immediate help within the first one hour of mountain rescue missions [206].

- **Cave Rescue**: The main objective of this type of SAR is rescuing survivors stuck in caves [88].
- **Urban Search and Rescue (USAR)**: USAR operations are usually about looking for victims in confined spaces [10]. It usually happens after the collapse of a structure. USAR has been studied extensively in robotics community [18].
- **Combat Search and Rescue (CSAR)**: CSAR involves rescuing soldiers during war or other military operations and transportation of them from a dangerous zone to a safe place [90]. There is usually a possibility of the presence of an enemy, which makes CSAR missions more challenging than other types of SAR missions [5].
- **Maritime Search and Rescue/Air-sea rescue**: As the name implies, search and rescue happens in maritime, i.e., in or near the sea [143]. Sometimes it involves searching for survivors of a sinking ship [170].
- **Wilderness Search and Rescue (WSAR)**: Here, the area of search covers quite large remote regions such as mountains, deserts, lakes, and rivers [71].

Search, rescue, and medical treatment teams form the key functional components of SAR teams [157]. SAR teams can also be categorized based on their operation and can be divided into two groups: command team(s) and field team(s) [92]. Depending on the country and SAR type, the SAR functional team structure can change significantly; however, the operational structure, i.e., the division to command team(s) and field team(s), remains the same: the command team is usually located close to the field. Their task is to manage the operation by allocating local tasks to field teams (search, rescue, medical treatment, etc.) efficiently, using all the global information they have available concerning the operation. Field teams, on the other hand, follow directions and commands given by the command team. They have local knowledge about their mission, and they need to stay in touch with the command team regularly to get necessary updates related to the operation. An example team organization for a rescue team were shown in Figure 2.1.

Regardless of the team organization of the rescue team or the type of SAR, time is critical for the success of missions [1]. Even in conventional rescue teams (without any robot), the interaction between team members directly affects the mission's outcome. Therefore, fast and efficient communication among rescue team members can be critical for saving

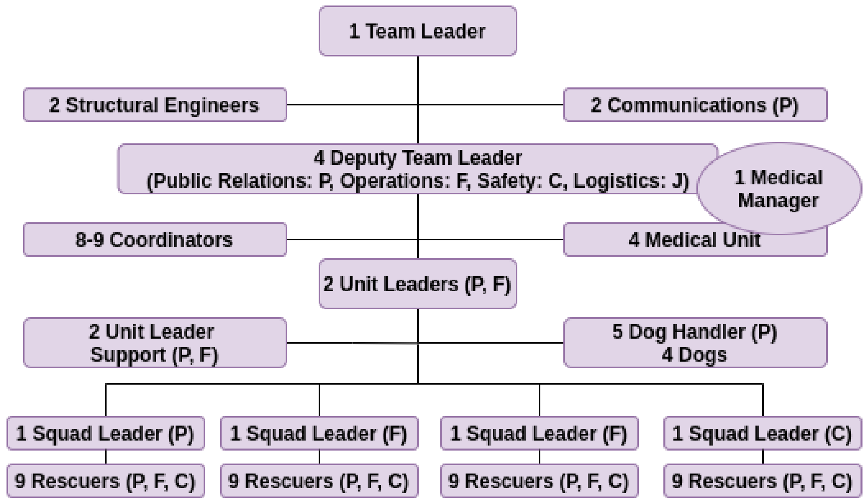


Figure 2.1: An example team organization of Japan Disaster Relief team who assisted to the Algeria Earthquake. The figure was drawn based on the original one in [158]. P refers to Police (NPA), F refers to Fire (FDMA), C refers to Coast (JCG) and J refers to Japan International Cooperation Agency (JICA)

people's lives, and bringing the usage of affective expressions into SAR context as studied in this thesis might a creative way to solve this problem.

2.1.1 Robots in Search and Rescue

The member composition of SAR teams has been changing over time. First, rescue dogs were included to help human SAR teams due to dogs' strong sense of smell that helps to find victims faster [190]. More recently, rescue robots became part of SAR teams. Various rescue robots have been successfully employed in real SAR operations depending on the SAR type such as snake robots [137, 85], shape-shifting robots [217], ground robots [20, 146], drones [166, 135], or underwater vehicles [147, 128]. There are many reasons behind the widespread use of rescue robots in real-life scenarios, such as (a) SAR areas are sometimes not safe or are unreachable for human rescuers due to various hazardous conditions such as extreme heat [28], the toxicity of the environment [73], or confined spaces [118], or (b) deploying robots to target SAR areas might require less time than deploying human rescue workers, and (c) the limited number of human rescue workers since training human rescue workers requires time and effort [2].

Research on robot-assisted SAR accelerated after the first time that rescue robots were utilized in the early 2000s by the Center for Robot-Assisted Search and Rescue (CRASAR) for the World Trade Center (WTC) disaster [144]. Initial research on SAR robots focused on the control of the robots, i.e., designing efficient and robust controllers to allow users to operate rescue robots easily [121]. Rescue robots equipped with such controllers have been used successfully during real-world applications such as searching for victims after the great eastern Japan earthquake [130].

After successful utilization of SAR robots, researchers shifted their focus to designing controllers or methods that can reduce human teleoperators' workload. First, low-level autonomous robot behaviours that were commonly used during SAR operations (e.g., the ability to climb up/down stairs autonomously without explicit human input [142]) were implemented. Then, since sometimes it is better to give human teleoperators a choice between full autonomy and manual teleoperation, semi-autonomous control methods with adjustable autonomy levels were implemented and tested in different scenarios, such as involving single robot-single operator [60] or single operator-multiple robot teams [211].

Machine learning (ML) techniques have been employed for robot-assisted SAR applications as well. First, they have been used to improve the efficiency of proposed controllers for SAR robots. For example, learning abilities were implemented through Hierarchical Reinforcement Learning (HRL) for semi-autonomous controllers of rescue robots used in USAR applications [52]. Later, researchers started to take advantage of ML methods to process sensory data that allowed SAR robots to understand the rescue environments better. In [153], deep reinforcement learning methods were used to help exploration of cluttered environments during USAR missions, which were found to be more efficient than conventional exploration algorithms. In another study [122], researchers employed deep learning techniques to detect open water swimmers in real-time (on-board) using GPS and camera data of drones, and they obtained 67% mean average precision (mAP), which is a high precision score for real-time onboard detection. These developments can help rescue workers to find victims faster.

Despite all the success in ML applications, some researchers started to question the black-box nature of the existing ML and deep learning methods and advocated eXplainable Artificial Intelligence (XAI) that encourages transparency and trustworthiness [7]. It is a general term that covers autonomous agents and robots as well as general AI concepts. From a human-robot interaction perspective, it is a well-studied finding that humans attribute mental states to agents or robots they are interacting with [80]. Hence, explainability of agents are needed to foster a natural interaction; otherwise, human users might (a) not trust the robot when it takes a correct action but does not justify it, thinking that the robot's action might be due to an error, or (b) assume that there is a logic behind every observed behaviour of a robot while there may not be a clear logic and an action may rather be a result of an internal error in the robot's decision-making system. In either case, the quality of interaction between humans and robots will get affected negatively. As a result, teamwork efficiency will be reduced significantly, which can lead to a severe reduction in users' trust in robot teammates [154]. To overcome these challenges, mental model architectures based on robots' sensory data were designed to observe the environment better and update robots' mental models about the environments [33]. In another study, researchers developed an explainable task planner that decomposes complex tasks into a sequence of transparent, simple tasks for unmanned aerial vehicles (UAVs). Such a planner was found to be easier for the human operator to understand and control compared to a planner that did not break down complex tasks into simple tasks [34].

While there has been a lot of progress in research related to robot-assisted SAR, rescue robots used today are still not as intelligent as human rescue workers since they cannot operate fully autonomously in cluttered real-life environments without human help [49]. Hence, it is safe to assume that humans will continue to work along with rescue robots. Research focusing on improving the interaction and communication between human and robot teammates is essential to guarantee the success of robot-assisted SAR.

2.1.2 Human-Robot Interaction in Search and Rescue

Most of the research in the HRI field related to robot-assisted SAR aims to improve teleoperation of SAR robots rather than focus on the interaction itself in human-robot teams [105]. In one study, researchers analyzed the trade-off between the number of human operators and the number of rescue robots in a team for the Robocup rescue competition, taking operators' decision time and mental workload as optimization parameters [181], and they found out that collaboration increases when the ratio of the number of operators over the number of robots increases for small-sized disaster areas. In another study, a specific simulation environment for USAR (USARSim) was employed to reduce human operators' mental workload and stress levels [116]. Operators who were assigned to a particular robot instead of selecting ones requiring attention from a shared pool reported a higher workload in the NASA-TLX questionnaire. In addition, some studies investigated swarm robots for SAR applications, but their focus was still on how to reduce human teammates' cognitive load, see e.g., [106]. A few studies address interactions between human and robot teammates in SAR. For example, in [105], RFID tags in the SAR environment were used to exchange information between teammates to increase mapping quality for better team performance. Their mapping method allowed teammates to focus on their task without worrying about changing their routes to get better map coverage. Researchers in another study developed a virtual reality simulation to simulate verbal communication in human multi-robot SAR teams, and they recorded data to create a better swarm emergency response where robots can clearly communicate with humans in the disaster area [32]. Researchers Hada and Takizawa [74] were successful in remotely controlling rescue robots from a long distance (700m) using ad-hoc radio signals. Although it was not implemented, usage of gestures to communicate with search and rescue UAVs was proposed in [131].

The work focusing on the social side of HRI in robot-assisted SAR is quite limited. In their research, Fincannon et al. found out that rescue workers expect SAR robots to have social capabilities [59]. Moreover, Murphy et al. conducted a survey with 28 medical doctors and therapists who operated rescue robots to interact with victims [148]. They discovered that it is vital for rescue robots to have social capabilities to relieve victims until physical assistance arrives. They also stated that having social intelligence may contribute to building less "creepy" rescue robots. This thesis explains an affordable way to express emotions for appearance-constrained rescue robots that is a helpful way to create more social (and less "creepy") rescue robots.

2.2 Affective Expressions in Human-Robot Interaction

Although integrating emotions into SAR robots have not been extensively studied by the HRI community so far, emotions, in general, have been one of the popular topics in HRI. HRI researchers have focused on how to use the embodiment of robots to express human-like emotions. In one study, researchers designed affective flight trajectories for drones to create expressive emotions for human users, taking inspiration from a performing arts method called the Laban Effort System [187]. In addition to employing motions to implement affective expressions, researchers also used color as an emotion indicator used for simple robots like Roomba [171] to humanoid robots like NAO [98]. Others employed sound as a way to communicate expressive emotions, but most often, motion, color, and sound were combined to create a better expressive system as in [193, 98]. Even sense of touch was used as an additional interaction modality for better expression of emotions in some studies. For example, in one study, researchers analyzed how human participants used their sense of touch to convey their emotions to a NAO robot [6]. In another study, sense of touch was used to create an affective robot to make participants calmer and happier [184]. There are also many studies focusing mainly on the utilization of facial expressions to implement expressive emotions for humanoid robots such as Kismet [22] and iCub [37], robots with a human-like face like MASHI [40], or robots with an animal-like face like Probo [178].

Despite all the work about implementing affective expressions for social robots, to the best of our knowledge, only one study attempts to use affective expressions on rescue robots [20]. In their research, Bethel and Murphy suggested design guidelines to use body movements, postures, orientation, color, and sound to implement non-facial and non-verbal affective expressions on the SAR robots iRobot Packbot Scout and Inuktun Extreme-VGTV and simulated a disaster site to conduct a user study to test the effectiveness of those suggested guidelines [20]. While these guidelines were used to create a social robot (which was compared with a robot that did not have these capabilities), they did not implement different emotions for the robot as we suggest in our research. Furthermore, unlike their work, we propose to use affective expressions as a complementary modality to increase the efficiency of multi-modal human-robot communication in SAR teams. We believe such an approach can provide further inside into SAR robotics that requires interdisciplinary work.

2.3 Emotions and Affect Control Theory

There have been many debates about the true nature of emotions since Darwin published "The Expression of Emotions in Man and Animals" in 1872 [43]. There are many different theories that define emotions, such as Ekman's Psychoevolutionary Theory of Emotions [55], James-Lange Theory [89] or Cannon-Bard Thalamic Theory of Emotions [27, 180]. In this paper, we are focusing on Ekman's definition [55]. According to him, emotions are caused by a specific event. He argues that basic emotions (sadness, happiness, fear, surprise, disgust, anger) are innate, present from birth, and universally recognized. Darwin also agreed on the universality of emotions and claimed that even people in isolated areas have similar emotional expressions [43].

Emotion prediction and modeling have been studied extensively by different research communities so far. Theories were introduced to divide emotions into different dimensions. Two well-known examples of these theories are the PAD emotional state model [134] and ACT [78]. These models use three dimensions: Pleasure, Arousal, Dominance (PAD) or Evaluation, Potency, Activity (EPA) dimensions, respectively, to describe the emotional meaning of words. Such dimensional emotional models usually have mappings that consist of ratings for different words (gathered through large surveys).

As these three dimensions are very similar in the two models (PAD emotional state model versus ACT), we decided to use the EPA dimensions (i.e., Evaluation, Potency, and Activity) of ACT in this thesis to implement affective expressions of Husky with lights quantitatively (Chapter 5), as there exist large mappings between emotions and EPA dimensions, which were gathered through large surveys and have been being updated over the years to account for possible changes over time.

The evaluation dimension in ACT shows how "good" an emotion, identity, action, etc. is (as rated through large surveys, the results of which suggests that there is an agreement in such ratings). The potency dimension shows how "powerful" something is, and the activity dimension shows how "active" it is. For example, the EPA value for the emotion "happy" is [3.44 ,2.93, 0.92][1], based on the U.S.A. 2015 Dataset [191] suggesting that "happy" is quite good, somehow powerful, and slightly active.

Before implementing emotions for SAR robots like the study explained in Chapter 5, analyzing whether there is a consensus in perception and expression of such emotions is necessary to verify whether communication through emotions would be effective; otherwise it will not be clear what emotion a robot should show in each situation, and there may be a risk of miscommunication. This is a crucial reason lying behind the first experiment presented in Chapter 3. In addition, finding a way to map SAR related situations to emotions is essential for utilizing affective expressions as an additional communication modality, which is the primary motivation of this thesis. To get a better way to match emotions with SAR scenarios, we took inspiration from the work in the sentiment analysis field and used ACT while conducting the second experiment presented in Chapter 4.

2.4 Sentiment Analysis

Our work in this thesis has similarities to research in sentiment analysis, namely, considering mapping sentences with emotions, moods, or sentiments. In general, sentiment analysis tries to extract emotions from sentences [201]. The classification of emotions in this process can be binary as in [204], where researchers categorized sentences as recommended (thumbs up) or not recommended (thumbs down) using unsupervised learning and obtained 74% mean accuracy. Some of the work in this area goes beyond getting only the mapping between sentences and emotions but tries to find the reason behind the predicted emotion (i.e., emotion stimuli). In [67], researchers trained a model to detect the best-associated emotion and its stimuli for given sentences among Ekman's six basic emotions (happiness, sadness, surprise, disgust, anger, and fear) [55] plus shame. After social media use became widespread, sentiment analysis methods were applied to sentences taken from social media posts. For instance, an algorithm called SentiStrength 2 was developed that detects sentiment strength of text from six social websites (MySpace, Twitter, YouTube,

[1] Note that EPA values are commonly rated in a range between -4.3 and 4.3

Digg, RunnersWorld, and BBC Forums) [201]. The algorithm performed better than a baseline approach for different datasets with both supervised and unsupervised methods using sixty-four standard emojis and 1246 million tweets. Since social media is used widely in different countries, other research focused on sentiment analysis in languages other than English such as Spanish as in [91]. Sentiment analysis research is also used to improve user experience in HCI field. For example, Setchi and Asikhia [186], did sentiment analysis to identify better image schema to provide an enjoyable user experience to users.

There is some work on the intersection of sentiment analysis and HRI. In [176], researchers took advantage of speech-to-text technologies to apply sentiment analysis on the conversation between humanoid robot MU-L8 and people interacting with it in order to improve the human-robot conversation. Researchers in [138], applied sentiment analysis methods to the feedback of customers interacting with the humanoid social robot Nadine to gain more insight about customers' expectations and how to use robots in real-world workplaces. Unfortunately, despite all the success obtained so far, the biggest limitation in this field is the fact that results highly depend on the context [150]. In other words, obtained mappings between text and emotions might differ drastically if the context of sentences changes. Hence, applying an approach similar to sentiment analysis to a specific context, namely SAR, as in the second experiment (Chapter 4) could contribute to research in SAR.

Chapter 3

Study 1: Matching Urban Search and Rescue Scenarios with Emotions

3.1 Research Questions and Hypotheses

This study aims to address the following research questions based on the discussion in the sections 2.1.2 and 2.2.

RQ1 Is there a consensus on what emotions to be used by USAR robots when they try to convey information about the situations commonly occurring during USAR operations?

RQ2 Is the mapping between emotions and USAR situations robust and not dependent on the wording of the sentences?

3.2 Methodology

In this experiment, participants first read about different USAR situations and messages and then were asked to map them to one or multiple emotions. To ensure that the wording of the messages would not affect the mappings, participants were divided into two conditions with two different communication styles. Depending on the condition, the messages were conveyed in either system status report style (e.g., "Dangerous material detected

Table 3.1: Situations/Messages that rescue robot(s) wants to convey in Experiment 1

No.	Condition A Social and Intelligent Conversational Agent Style	Condition B System Status Report Style
(1)	I can again communicate with our team outside of the building	Communication with external team restored
(2)	I lost communication with our team outside of the building	Communication with external team lost
(3)	I am stuck and might need help to proceed	Stuck here
(4)	I detected dangerous material here, let's proceed carefully	Dangerous material detected here
(5)	I believe we are behind schedule. I also noticed it is getting dark and there is not much time left	Behind schedule. It is getting dark
(6)	I found an item that could belong to a person. Maybe the person is nearby	An object that might belong to a person was found
(7)	My battery is running low and I will stop working soon	Battery is running low
(8)	I think I found a surviving person	Possible living person detected
(9)	I detected that there might be a risk of further collapse so we should only proceed with caution	Further risk of collapse detected
(10)	I think I heard someone is calling for help, we might have found a survivor	Possible call for help detected

here") or social and intelligent conversational agent style (e.g., "I detected dangerous material here, let's proceed carefully"). Note that the social, conversational agent style also had additional information (e.g., "let's proceed carefully") that is intuitive and does not exist in the other style. The messages in each condition were not meant to be identical in content. If significant differences are observed in how participants map a robot's communication with emotions in these two conditions, this would indicate that a mapping will depend heavily on the communications style, which could make such a mapping less effective. On the other hand, if we find similar mappings in both conditions, this would suggest that such a mapping can be used across different robot communication styles. The messages are shown in Table 3.1.

For each message, participants could choose one or multiple options from a list of 11 affective labels (including emotions as well as moods, but in the context of this article referred to summarily as 'emotions'): Bored, Sad, Surprise, Calm, Disgust, Angry, Tired, Annoyed, Fear, Happy, and Excited. We focused on the affective interpretation of the messages rather than how the emotions themselves can be communicated. Providing participants with an emotion list as a choice also helps to avoid ambiguity which might be the case if we used another way of presenting emotions such as emoticons [50]. This set was suggested by Ghafurian et al. [66] and it covers a wide range of affective expressions, including simple and complex emotions, as well as different moods. Note, the existing design guidelines [66] on how to implement these emotions for zoomorphic robots could in the future facilitate implementing them on a zoomorphic USAR robot.

All situations shown in Table 3.1 were selected carefully after analyzing common situations occurring during SAR missions. For example, the command team, which is considered to be the central control unit, is responsible for gathering all the global information and sharing it with the field teams. There are regions where the field teams cannot make a contact with the command team, which are called dead zones. It is quite critical for field workers to be aware of whether they are in the dead zone or not [92]. That is why situations (1) and (2) were included.

Introducing robot teammates creates some problems for field workers as well. They need to know whenever the robot teammate(s) need help. Common problems that rescue robots have are being stuck in an area and having low battery [49]. These problems were included as situations (3) and (7). Information about these situations would help to increase the transparency of robot teammates' actions [111].

Situation (5) was included as timing is critical in SAR missions, and field workers need to know and take active or proactive actions when their performance slows down [156]. The probabilities of further collapse in the searched area [159] or encountering hazardous material [73] are high for USAR missions. Therefore, situations (4) and (9) were included. Situations (6), (8), and (10) were considered since they represent the most positive scenarios that might happen during a SAR mission and all the field workers need to be aware of these situations whenever they happen [92]. Lastly, an additional message was added as an attention check, where we instructed participants to 'select happy as the answer'.

Different communication styles used for the statements were inspired by research in HCI and HRI, where agents/robots are attributed human-like traits based on their behaviour [65]. Their communication may differ and be closer to humans (human-like) or machines (machine-like) [81, 197, 173] (i.e., social and intelligent conversational agent style and system status report style, respectively, in our case).

Furthermore, participants were asked to respond to a questionnaire. The questionnaire was composed of two sections: demographic questions (i.e., age, gender, education level, and ethnicity) and questions related to emotions and SAR experience, which contained the following questions (all were answered on a continuous Likert scale from completely disagree to completely agree, advantages of which are discussed in [203, 127, 62])[1]:

- I think rescue robots are useful. (Robots-useful)
- I was familiar with rescue robots before this study. (Familiar-SAR)
- I think rescuing people can save their lives. (Attention check question)
- I had seen an example of rescue robot before this study. (Seen-SAR-Robots)

[1]note that some questions were repeated in different ways, which acted as sanity checks

- I think rescue robots are necessary. (Robots-necessary)
- I believe in future rescue robots can become better than rescue dogs.
- I think rescue robots are not useful.
- I believe in the future we will not need rescue robots.
- I am good at showing proper emotions in real life situations. (Good-show-emotions)
- I have difficulty understanding others' emotions.
- I have difficulty showing proper emotions.
- I think when people are happy, they mostly express happy emotions. (Attention check question)
- I think I have a good understanding of people's emotional states/moods. (Good-understand-emotions)

3.3 Experiment 1

Through a user study on Amazon Mechanical Turk, we investigated whether a natural and consistent mapping exists between a robot's communications/messages in USAR missions and emotions.

3.3.1 Procedures and Measures

Participants first accepted the consent form and read the instructions. They then read a short example of an USAR scenario. They saw various images of rescue robots to better understand the concept (and also to ensure that their assumption of the robot would not affect their responses). In order to give participants examples of rescue robots and prime them towards considering USAR scenarios in general and communication situations with USAR robots in particular, we showed a variety of different existing USAR robot designs. But to limit the information provided, trying to avoid the elicitation of answers regarding any detailed features of specific robots, we presented five modified example pictures of the existing USAR robots with different shapes, e.g., human-like, machine-like, animal-like, etc. For this purpose, we altered the original pictures into black and white line drawings.

Figure 3.1 shows an example of a machine-like USAR robot shown to participants. Since the focus of the study is about how to associate messages sent by the robot with emotions, there are no visible speakers or other mechanisms to explain how the robot sends the message to participants. Instead, they were told that the robot wants to communicate a specific message, as can be seen in Figure 3.2.

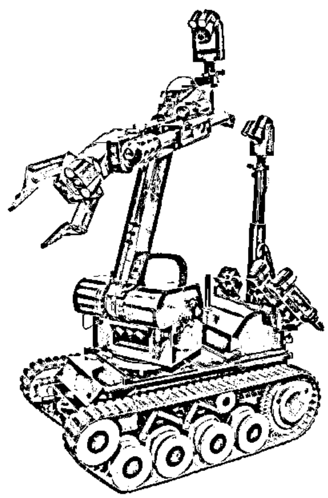


Figure 3.1: An example of the USAR robot illustrations used in the study. The original picture was adapted from [47]. Used with permission. Modifications by us resulted in a black and white line drawing.

After reading the example USAR scenario and getting familiar with the concept, participants saw the statements in Table 3.1 in a random order, and they were asked to select one or multiple emotions that they thought would be appropriate for a robot to show in that situation. They were specifically instructed to choose an emotion, or emotions, that they thought a rescue robot could show to convey particular situations to rescue workers (see Figure 3.2 for an example). After the completion of mapping all ten sentences to emotions, participants answered the questionnaire.

Since this study was exploratory, we did not administer many pre/post-test questionnaires. Selected emotions corresponding to specific messages/situations gave the mapping we were after. However, we asked participants demographic questions, their familiarity with search and rescue operations, and their opinion about rescue robots to analyze possible correlations between obtained mappings and participants' demographic info and/or their personal views about the subject. During the task, the participants were asked one attention check question related to the task to make sure that they were paying attention to the scenarios. This is a standard method with online crowd-sourced studies to improve the quality of obtained data [103].

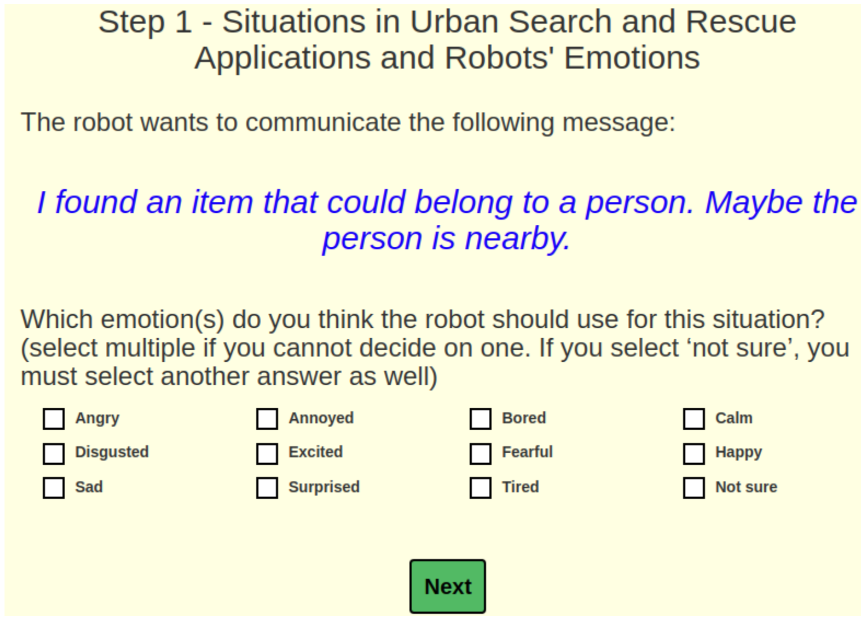


Figure 3.2: Interface Used in Experiment 1

3.3.2 Participants

We recruited a total of 112 participants on Amazon Mechanical Turk. All participants had an approval rate higher than 97% based on at least 100 HITS. Fifteen participants failed the attention check questions, and 19 participants provided inconsistent answers to the questionnaire at the end of the study. This left 78 participants (48 male, 29 female, 1 other; ages 20-72, avg: 35.7). Out of the remaining 78, 40 saw messages in the social and intelligent conversational agent style, and 38 saw them in the system status report style. Participants were paid $2 if they completed the task. Otherwise, they were paid a pro-rated amount based on the number of questions they completed. This study received ethics clearance from the University of Waterloo's Research Ethics Committee (see Appendix A).

3.3.3 Results

In this section, we present the obtained results related to mappings of statements to emotions for the two different styles of statements, as well as the questionnaire results.

Communication Style Effect

We first checked how different wording of the sentences (i.e., social intelligent conversational agent style vs. system status report style) affected the responses. All pairs of corresponding sentences in each condition were significantly correlated ($0.78 \leq r \leq 0.99$), suggesting that the selected mappings were robust and wording did not affect the chosen emotion.

Mapping Results

Figure 3.3 and Table 3.2 shows the results for the selected emotions for each sentence and each wording style. Each cell in this table (except for the cells in the last two columns) shows how many participants chose the corresponding emotion. Significance of the selected responses was calculated using one-way binomial tests, assuming uniform probability distribution as the null hypothesis (i.e. to measure whether options are selected significantly more than the chance level), and significant responses were shown in bold with corresponding significance levels represented in Figure 3.3 and Table 3.2 (see [178] for an example of a related study and statistical analysis). Furthermore, one-way binomial tests were also employed to check whether a specific response was selected significantly more than all the other options for a specific emotion (shown with pink). For example, happy was selected significantly more than the other emotions for the statement "Communication with external team restored".

In some cases, we could not find a specific emotion that was selected significantly more than others. Instead, we found a set of 2 or 3 emotions that were selected considerably more than random. For example, in the statement "Possible living person detected." both happy and excited were preferred, i.e., they were selected significantly more often than random. We also included the mean and standard deviation (std) for the number of choices selected for each statement, which provide information on the number of emotions that the participants selected and found to be a good mapping for a particular message. Both mean and std were very low for all statements, indicating a higher certainty in the choices. Overall, the results showed that the participants had preferences for specific emotions for each message and that such mapping is significant and consistent.

Table 3.3 summarizes the results shown in Table 3.2 and Figure 3.3. It contains the set of emotions that were selected significantly more than random for each statement. Our suggestions for the final mappings are shown in green, and possible alternative mappings are shown in orange.

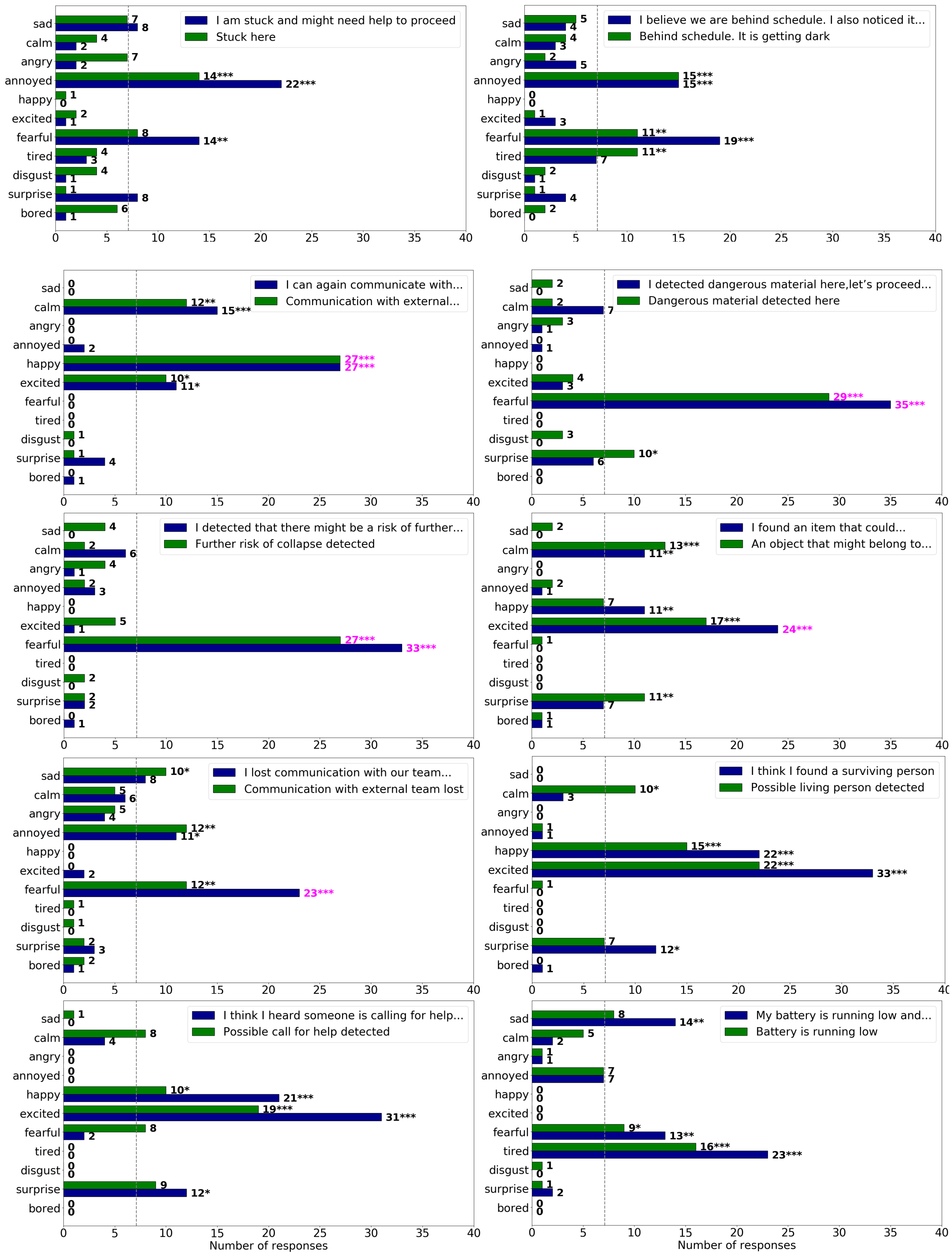


Figure 3.3: Each bar graph shows the total number of selected emotions for each sentence and condition. Stars represent significance levels that were calculated via binomial tests assuming random selection as a null hypothesis (***: $p < .001$, **: $p < .01$, and *: $p < .05$). Emotions that were selected significantly more than the others are shown in pink (all pairs selected significantly more than the chance was compared). Gray dashed lines show the chance lines.

Table 3.2: Number of responses for each emotion and statement for both wording styles. Emotions selected significantly more than random are shown: ***: $p < .001$, **: $p < .01$, and *: $p < .05$ (significance was calculated through binomial tests). Emotions selected significantly more than all other options for each statement are shown in pink. *Number of selections* shows mean and standard deviation for the number of selected responses for each statement.

statement	cond.	bored	surprise	disgust	tired	fearful	excited	happy	annoyed	angry	calm	sad	# of Selections mean	# of Selections std
I am stuck and might need help to proceed.	A	1	8	1	3	**14****	1	0	**22*****	2	2	8	1.55	0.78
Stuck here.	B	6	1	4	4	8	2	1	**14*****	7	4	7	1.53	0.65
I believe we are behind schedule. I also noticed it is getting dark and there is not much time left.	A	0	4	1	7	**19*****	3	0	**15*****	5	3	4	1.52	0.78
Behind schedule. It is getting dark.	B	2	1	2	**11****	**11****	1	0	**15*****	2	4	5	1.45	0.60
I can again communicate with our team outside of the building.	A	1	4	0	0	0	**11***	**27*****	2	0	**15*****	0	1.52	0.68
Communication with external team restored.	B	0	1	1	0	0	**10***	**27*****	0	0	**12****	0	1.39	0.55
I detected dangerous material here, let's proceed carefully.	A	0	6	0	0	**35*****	3	0	1	1	7	0	1.32	0.57
Dangerous material detected here.	B	0	**10***	3	0	**29*****	4	0	0	3	2	2	1.39	0.64
I detected that there might be a risk of further collapse so we should only proceed with caution.	A	1	2	0	0	**33*****	1	0	3	1	6	0	1.22	0.53
Further risk of collapse detected.	B	0	2	2	0	**27*****	5	0	2	4	2	4	1.26	0.55
I found an item that could belong to a person. Maybe the person is nearby.	A	1	7	0	0	0	**24*****	**11****	1	0	**11****	0	1.37	0.54
An object that might belong to a person was found.	B	1	**11****	0	0	1	**17*****	7	2	0	**13*****	2	1.45	0.64
I lost communication with our team outside of the building so we are on our own now.	A	1	3	0	0	**23*****	2	0	**11***	4	6	8	1.47	0.64
Communication with external team lost.	B	2	2	1	1	**12****	0	0	**12****	5	5	**10***	1.34	0.48
I think I found a surviving person.	A	1	**12***	0	0	0	**33*****	**22*****	1	0	3	0	1.8	0.65
Possible living person detected.	B	0	7	0	0	1	**22*****	**15*****	1	0	**10***	0	1.5	0.76
I think I heard someone is calling for help, we might have found a survivor.	A	0	**12***	0	0	2	**31*****	**21*****	0	0	4	0	1.75	0.81
Possible call for help detected.	B	0	9	0	0	8	**19*****	**10***	0	0	8	1	1.47	0.65
My battery is running low and I will stop working soon.	A	0	2	0	**23*****	**13****	0	0	7	1	2	**14****	1.55	0.68
Battery is running low.	B	0	1	1	**16*****	**9***	0	0	7	1	5	8	1.34	0.53

Table 3.3: Statements and the affective expressions that were selected significantly more than random (***: $p < .001$, **: $p < .01$, and *: $p < .05$). Our suggested first choice for the mapping is shown in Green (which is also consistent with the ones shown as pink in Figure 3.3 and Table 3.2). The second, alternative, suggestion is shown in Orange. Cond. stands for condition.

Message	Cond.	Affective Expression
I am stuck and might need help to proceed.	A	fear**, **annoyed*** **
Stuck here.	B	**annoyed*** **
I believe we are behind schedule. I also noticed it is getting dark and there is not much time left.	A	**fear*****, **annoyed*** **
Behind schedule. It is getting dark.	B	tired**, **fear****, **annoyed*** **
I can again communicate with our team outside of the building.	A	excited*, **happy*****, **calm*** **
Communication with external team restored.	B	excited*, **happy*****, **calm****
I detected dangerous material here, let's proceed carefully.	A	**fear*** **
Dangerous material detected here.	B	surprise*, **fear*** **
I detected that there might be a risk of further collapse so we should only proceed with caution.	A	**fear*** **
Further risk of collapse detected.	B	**fear*** **
I found an item that could belong to a person. Maybe the person is nearby.	A	happy**, **excited*****, **calm****
An object that might belong to a person was found.	B	surprise**, **excited*****, **calm*** **
I lost communication with our team outside of the building so we are on our own now.	A	**fear*****, **annoyed***
Communication with external team lost.	B	sad*, **fear****, **annoyed****
I think I found a surviving person.	A	surprise*, **excited*****, **happy*** **
Possible living person detected.	B	calm*, **excited*****, **happy*** **
I think I heard someone is calling for help, we might have found a survivor.	A	surprise*, **excited*****, **happy*** **
Possible call for help detected.	B	**excited*****, **happy***
My battery is running low and I will stop working soon.	A	sad**, **tired*****, **fear****
Battery is running low.	B	**tired*****, **fear***

Questionnaire Results

As mentioned above, we asked participants about their familiarity with USAR situations and robots used in USAR, as well as their perception of necessity and usefulness of robots (to understand their general attitude toward robots). Participants were also asked to self-evaluate their ability to understand and show emotions. This was mostly to understand our participant group better and to identify if any of these factors have affected ratings. Figure 3.4 shows the obtained results. The majority of the participants believed that the robots are necessary and useful in USAR situations and indicated that they are good in understanding and showing emotions. However, the majority of participants were not familiar with USAR scenarios, neither had they seen a USAR robot before. We did not find any effect of these factors on participants' mappings.

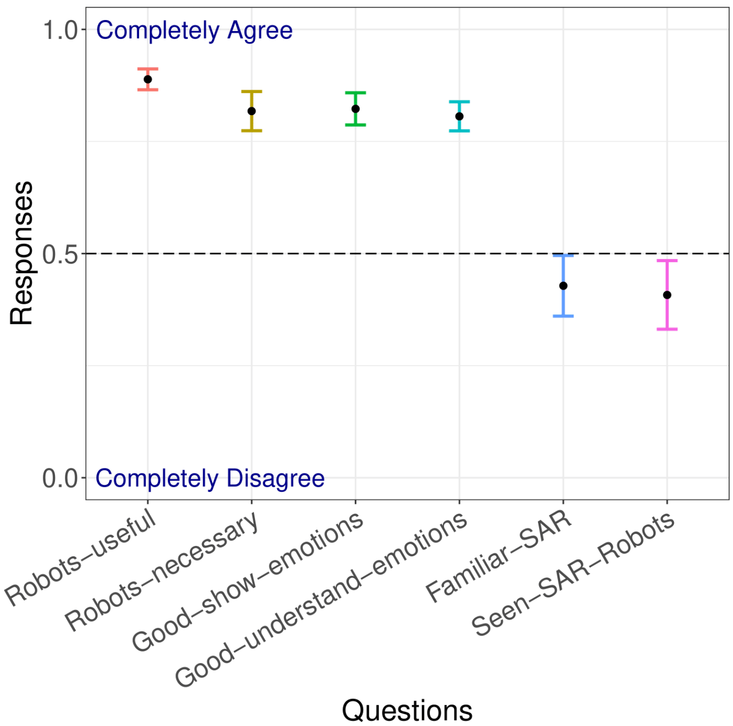


Figure 3.4: Participants' responses to the questionnaire in experiment 1

3.3.4 Discussion

Using emotions as an additional communication channel for multi-modal interaction can improve interactions between humans and robots, as perceiving emotions are considered to be natural for people [115]. In this experiment, we investigated the feasibility of using emotions in USAR for conveying a robot's messages. Through a study on Mechanical Turk, mappings between 10 common situations in USAR and the associated affective expressions (11 options in total) were investigated.

Regarding **RQ1**, the results suggested that reaching consensus on specific mappings between messages in USAR situations and emotions is feasible, as the majority of the participants agreed on the same emotions. Further, related to **RQ2**, the mappings seemed to be robust, as they were not affected by the wording of the sentences, i.e., the robot's communication style. In other words, conveying statements in social and intelligent conversational agent style or system status report style did not affect the selected mappings between emotions and statements.

Another interesting point to discuss the obtained mapping is the range of selected emotions. Despite providing 11 different emotions/affective states to the participants, three of these emotions (bored, disgusted, and angry) were not preferred by the majority for any message. We can only speculate at this point, but the reason behind this might be the fact that the total number of statements during the study was limited, and participants could not find a proper sentence to match these emotions. This suggests that a subset of emotions might be sufficient to convey information from robots to humans, as they will be complement other types of multi-modal communication. Furthermore, while 'happy' and 'excited' were selected together in many situations, we noticed a preference towards one of them depending on the statement.

It is essential to highlight that we propose to use emotions as an additional modality in the specific USAR context to convey information from robots to humans, complementing the current multi-modal communication methods (rather than replacing them). As a result, robots will be able to communicate with human teammates through emotions in addition to current modalities such as voice, text, and videos. Using this additional modality of emotions and affective expressions could help with designing more robust and fail-safe human-robot communication systems, which is particularly crucial in high-risk real-life applications, such as USAR or firefighting. Furthermore, including emotions as a modality in the USAR context can provide a shared communication model between different cultures and can reduce stress levels of victims [20].

When using emotions as a modality, there might be a risk of clouding the communication with emotion signals. While future work is needed to study if this can affect communication in USAR situations, and find out how to minimize this risk, it is expected that as the affective expressions are intuitive to interpret, this risk is low. USAR robots' capability to detect/predict field workers' emotional states, and to consider those while generating appropriate affective expressions, could also further improve human-robot interaction and communication and help reduce this risk.

Ultimately, the idea of using emotions as an additional communication modality to convey information about robots' internal state could be generalizable to various types of robots. Here, we did not discuss how emotions will be expressed intentionally, as we first needed to understand which emotions would be considered suitable (regardless of the communication method) for each situation. Although the way robots express the emotions might change, the mapping between emotions and messages might stay the same for the USAR context. Ultimately, the communication method would depend on the robot, and different communication methods (and a combination of them) need to be evaluated to find the most suitable approach. In addition, one can employ the same idea to create an intuitive alternative communication channel for other application scenarios that require human-robot teamwork, such as firefighting, peacekeeping activities, warehouse applications.

3.4 Limitations and Future Work

This study had several limitations. First, while we limited participation to North America, we did not check participants' English proficiency. Given the geographical constraint for recruitment and the consistency in responses, it is reasonable to expect that most participants clearly understood the statements and questions. Second, many of our participants were not familiar with USAR scenarios or the robots used in those scenarios (see Figure 3.4). They were also asked to imagine the situations, as opposed to interacting with a real robot. While we tried to reduce this effect by showing participants examples of such robots and providing information about USAR, their responses might have been affected. Future work needs to investigate these mappings with the participation of search and rescue workers through an online study and then in real USAR situations where rescuers are interacting with actual robots. In the present study, we did not investigate real interactions to not bias participants' responses by choosing a specific robot and its specific appearance and behaviours. Once the mappings are validated and robust, it will be a natural next step to test them in real scenarios.

Third, while we selected a subset of affective expressions that covered the primary emotions and other, more complex emotions and moods, this set's choice might have affected participants' responses and different mappings might have been obtained with other sets. Also, only a small group of emotions were selected by participants at the end. Future study is needed to analyze the efficiency of the smaller subset further.

Future work is also needed to investigate the implementation of these emotions in robots. Although we have validated the perception of affective expressions in another study [66] on a social, zoomorphic robot called Miro [39], we need further studies to implement these affective expressions on non-zoomorphic robots that are already used in USAR operations and for which the existing design guidelines (e.g., as provided by Bethel and Murphy [20]) need to be considered.

Note, we only considered two communication styles for the robot. Future work could further test the robustness of our mapping by also considering other communication styles.

Additional studies using the idea of mapping between specific messages and emotions for different application scenarios could also solidify the generalizability of the concept of conveying information via emotions as a complementary communication modality in USAR.

Lastly, future work can consider the applicability of this approach to other related application areas, such as using human-robot teams in firefighting.

3.5 Conclusion

- This experiment was an exploratory study to investigate the feasibility of using emotions as an additional communication modality in robot-assisted USAR.
- There was a consistency in participants' responses to match particular USAR related sentences with emotions resulted in one or two suggested mappings between common situations occurring in USAR and emotions that the robot should show in those situations.
- The mapping between emotions and USAR situations were robust to the wording of the situations (i.e., intelligent agent style versus system status report style).
- Findings of the experiment supported the idea of employing emotions to communicate messages in the USAR context.

Chapter 4

Study 2: Affect Control Theory and the Mapping Between Search and Rescue Scenarios and Emotions

Since Experiment 1 suggested that mapping emotions to situations in SAR are feasible, we asked if there is a method to obtain these mappings in a way that (a) the mappings would not solely depend on a set of emotions (e.g., the 11 emotions shown to the participants in the previous study), and (b) the mapping process would have a potential to be automated in the future. Therefore, in this experiment, we study whether it is possible to use the three dimensions associated to emotions in the PAD emotional state model [134] (PAD) and ACT [78] (EPA), to address RQ3. As these three dimensions are very similar in the three models, we decided to use the EPA dimensions of ACT, as there exist large mappings between emotions and EPA dimensions, which were gathered through large surveys and have been being updated over the years to account for possible changes over time.

The evaluation dimension in ACT shows how "good" an emotion, identity, action, etc. is (as rated through large surveys, the results of which suggests that there is an agreement in such ratings). The potency dimension shows how "powerful" something is, and the activity dimension shows how "active" it is. For example, the EPA value for the emotion "happy" is [3.44 ,2.93, 0.92][1], based on the U.S.A. 2015 Dataset [191], which is used in this experiment, and which suggests that "happy" is quite good, somehow powerful, and slightly active.

[1]Note that EPA values are commonly rated in a range between -4.3 and 4.3

4.1 Research Questions and Hypotheses

This study aims to address the following research question based on the discussion in the sections 2.1.2 and 2.2.

RQ3 How can a mapping between SAR related sentences and emotions be obtained, and is there a way to generalize such mapping without limiting it to a specific set of emotions?

4.2 Methodology

To study whether using the EPA dimension can lead to similar mappings, instead of using a set of emotions (as in Experiment 1), we asked participants to rate the sentences used in Experiment 1 on the evaluation, potency, and activity dimensions (see Figure 4.1). Afterward, we calculated the emotion that is the closest to the EPA rating for each sentence which described in more detail later in this section.We also included more sentences related to different types of SAR in addition to the sentences pertaining to USAR in the first experiment to check the validity of such a mapping (i.e., if meaningful mappings can be obtained) for an extended set of situations. Since Experiment 1 suggested that the mappings were consistent and not affected by different wording styles, we only focused on sentences conveyed in social and intelligent conversational agent style, which is more expressive, for this study.

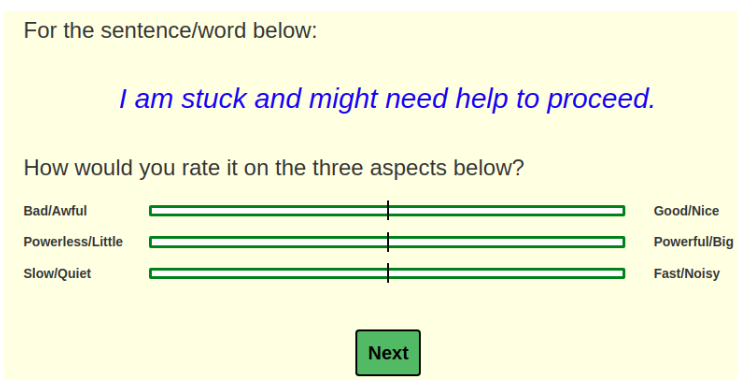


Figure 4.1: Interface Used in Experiment 2

Selection of additional sentences: Generally, the sentences used in the first experiment were quite inclusive. In other words, they can be used for other SAR types as well. Nonetheless, as we wanted to increase the set of sentences in this study to make it more comprehensive, we included a few other situations relevant to other types of SAR operations. Table 4.1 shows the extended set of sentences used in this study. Sentences (11) and (12) were included since there generally exists more than one field team in the search area, and the need for additional members might change dynamically depending on the given task [92]. Sentences (13) and (14) were included because detecting the environment's temperature is crucial for those SAR sub-types that involve extreme areas like deserts, water, or icy environments. That is because the survival rate of victims decreases significantly both in cold water during maritime search and rescue [170] and in hot weather due to dehydration during WSAR [84]. Furthermore, sentence (15) was added since there is a chance to encounter an injured victim in all types of SAR operations [92, 143, 5, 95]. Lastly, sentence (16) represents another scenario that is common during almost all SAR operations since it is usually impossible to reach some area of interest in the rescue directly field [88, 206, 71, 170].

Table 4.1: All Situations/Messages that rescue robot(s) wants to convey in Experiment 1 and 2

No.	Sentences	Study
(1)	I can again communicate with our team outside of the building	1,2
(2)	I lost communication with our team outside of the building so we are on our own now.	1,2
(3)	I am stuck and might need help to proceed	1,2
(4)	I detected dangerous material here, let's proceed carefully	1,2
(5)	I believe we are behind schedule. I also noticed it is getting dark and there is not much time left	1,2
(6)	I found an item that could belong to a person. Maybe the person is nearby	1,2
(7)	My battery is running low and I will stop working soon	1,2
(8)	I think I found a surviving person	1,2
(9)	I detected that there might be a risk of further collapse so we should only proceed with caution	1,2
(10)	I think I heard someone is calling for help, we might have found a survivor	1,2
(11)	I think we need additional team members	2
(12)	I think we have more team members than we need. One of us should join the other team	2
(13)	I detected that the temperature of the environment is too cold for a person	2
(14)	I detected that the temperature of the environment is too hot for a person	2
(15)	I found a victim that requires medical attention	2
(16)	I detected that this rescue route requires obstacle clearance	2
(17)	I found that for this sentence you have to select the leftmost option on all bars (attention check)	2
(18)	I detected that for this sentence you must select the rightmost option on all bars (attention check)	2

Rating the sentences: The design of the interface and questions in this experiment was similar to Experiment 1: participants saw all the sentences in Table 4.1 in a random order, and they were told to rate these sentences according to three dimensions: (a) evaluation, (b) potency and (c) activity, which are known as EPA in ACT [78]. In other words, participants were asked to rate, on a continuous scale, how good, how powerful, and how active each sentence (and the corresponding situation it conveys) was (see Figure 4.1).

We did not use the demographics or other questionnaires from Experiment 1 as we did not find a correlation between questionnaire results and the mappings obtained from Experiment 1. However, as a consistency check, we asked participants to rate words "angry", "good", "infant", and "boss" in addition to the sentences to compare these ratings with the original EPA values obtained from the U.S.A. 2015 Dataset [191]. These words were selected as they cover a range of different values on each of E, P, and A dimensions and could help ensure consistency between participants' rating and the ratings in the dataset, based on which the final mappings to the emotions were calculated. In addition to these sentences and words, attention checks were included that instructed participants to select a specific answer (e.g., "I found that for this sentence you have to select the leftmost option on all bars.").

Emotion	tired	surprise	sad	happy	fear	excited	disgust	calm	bored	annoyed	angry
(E)	-1.55	1.42	-2.29	3.44	-2.37	2.69	-2.27	2.88	-1.85	-2.08	-1.77
(P)	-1.28	1.35	-1.44	2.93	-1.04	2.18	0.22	1.93	-0.86	-0.57	0.57
(A)	-2.30	2.17	-2.04	0.92	-0.71	2.24	0.43	-2.32	-2.01	0.53	1.80

Table 4.2: The EPA values for the 11 emotions used in Experiment 1 based on the U.S.A. 2015 Dataset [191]

Obtaining the associated mappings: Obtained EPA ratings were used to identify the corresponding emotion among the list of 11 affective expressions used in the first study. The same set of 11 was used to ensure consistency. The EPA ratings of the 11 emotions were obtained from the U.S.A. 2015 Dataset [191] (See Table 4.2), which is the most recent dataset. Euclidean distance [42] was used to find the closest mapping. As an example, for the sentence "I think we need additional team members", we compared the distances between participants' EPA ratings for this sentence (e.g., [0.83,0.77,0.71]) and ratings of all 11 emotions, and we found that the closest distance (1.68) corresponded to the emotion "surprised" ([1.42,1.35,2.17]) as shown below:

```
#Read EPA dictionary with all words
EPA_dict = read_dict(USA_Surveyor_2015)

emotion_list = [bored, sad, surprise,
```

```
calm, disgust, angry, tired, annoyed,
fear, happy, excited]

#Select only EPA ratings in the list
EPA_dict = select(EPA_dict, emotion_list)

#Average EPA ratings for I think we need
#additional team members sentence
v1 = (0.83, 0.77, 0.71)

for each emotion in EPA_dict
    #Calculate distance between EPA
    #ratings of emotions and the sentence
    v2 = emotion[rating]
    dist = euc_dist(v1,v2)
    if dist < min_dist
        min_dist = dist
        predicted_emotion = emotion[name]

print(emotion[name], emotion[rating])
#surprised, (1.42, 1.35, 2.17)
print(min_dist)
#1.68
```

4.3 Experiment 2

4.3.1 Procedures and Measures

Participants first read the consent form and the instructions for the study. Afterwards, they rated sentences related to SAR (see Table 4.1) and the words "angry", "good", "infant", and "boss" on the EPA dimensions (see Figure 4.1), the order of which was randomized. Finally, they received an end code for the completion of the study.

4.3.2 Participants

We recruited 223 participants (79 from Canada and 144 from the USA) on Amazon Mechanical Turk for this study. We started with the same recruitment criteria as in Experiment 1:

having an approval rate of at least 97% based on at least 100 HITS on Mturk, but later changed this criterion to an approval rate of 96% based on at least 50 HITS on Mturk for participants who were from Canada to recruit more participants [2]. After filtering, based on the attention check questions (17) and (18) in Table 4.1, data from a total of 133 participants were left (72 from Canada and 61 from the USA). Participants were paid $0.30 for participation in this study. This study received ethics clearance from the University of Waterloo's Research Ethics Committee (see AppendixA).

4.3.3 Results

In this section, we will first discuss how consistency checks were applied, and will then present the results for the ratings and the obtained mapping between the situations and emotions.

Consistency Check

Despite having a high approval rate criteria for recruitment on MTurk, 90 participants failed either/both of the attention check questions. Since attention check questions were related to selecting the right or left-most part of the bars, we included an error margin during the filtering and accepted a range of answers that were not too far from the correct answer on the continuous scale (i.e., a 10% error margin for both the left and right most part of the continuous scale).

After removing those who failed our attention checks, the mean EPA values of participants who passed attention checks from both countries were compared to study whether/how cultural differences affected the ratings. As it can be seen in the correlation column of Table 4.4, we found high correlations between the ratings from the two countries, suggesting that the ratings were consistent between Canada and the USA. Therefore, the merged data is used for analyzing the results.

Scaling

As we used a specific EPA dataset to find the closest emotion to each of the sentences, we first had to ensure that participants' ratings are consistent with those in the dataset.

[2]None of the participants recruited from Canada failed any of the attention checks. Therefore we changed the criteria to be able to recruit more participants

Therefore, we first checked participants' ratings of words "angry", "good", "infant" and "boss". Averages of these EPA ratings were calculated and they were compared with EPA ratings obtained from the U.S.A. 2015 Dataset [191] using Pearson's correlation coefficient [16]. We found high correlations (see the last column of Table 4.3), which suggested that there would be no need to scale the obtained EPA ratings assessed by the participants, and we can perform the mapping on the results.

Table 4.3: Comparison of mean EPA ratings scored by participants with EPA ratings obtained from the U.S.A. 2015 Dataset [191]. Results of Pearson's correlation coefficient are shown in the last column.

Words	**From Participants**			**From Dictionary**			**Corr**
	E	**P**	**A**	**E**	**P**	**A**	
angry	-3.08	1.44	2.23	-1.77	0.57	1.80	0.98
good	3.69	2.32	0.45	3.40	2.37	-0.24	0.99
infant	2.62	-2.45	-0.43	2.26	-2.35	1.23	0.91
boss	0.77	3.07	1.61	0.91	2.79	1.07	0.96

Mapping Results

The results for EPA ratings, as well as the mapping outcomes are shown in Figure 4.2 and Table 4.4. Each row in Table 4.4 contains the mean EPA values for a particular sentence and the two closest predicted emotions, calculated through the above-mentioned method (i.e., by comparing Euclidean distances between mean EPA ratings and EPAs of the 11 emotions according to the U.S.A. 2015 Dataset [191]). For each of the predicted emotions, the calculated distance (dist.) is stated. We also show the results from Experiment 1 in the last column. For example, for the sentence "I can again communicate with our team outside of the building", participants' average EPA ratings were: $E = 2.55, P = 1.73, A = 0.93$ [3]. The correlation between the ratings in the US and Canada was 0.98, the closest emotion to average EPA ratings of participants was calculated to be "Excited" with a distance of 1.39, compared to EPA ratings in the dataset, and the second closest emotion was "Happy" with a distance of 1.49. These results were consistent with the mappings obtained from Experiment 1 (i.e., Excited, Happy, and Calm). Only for two of the sentences, the two closest emotions did not match with the ones obtained through the first experiment, but the third closest emotion matched. The third closest distance for these two sentences was shown in pink in Table 4.4.

[3] As a reminder, EPAs are commonly rated in a range between -4.3 and 4.3

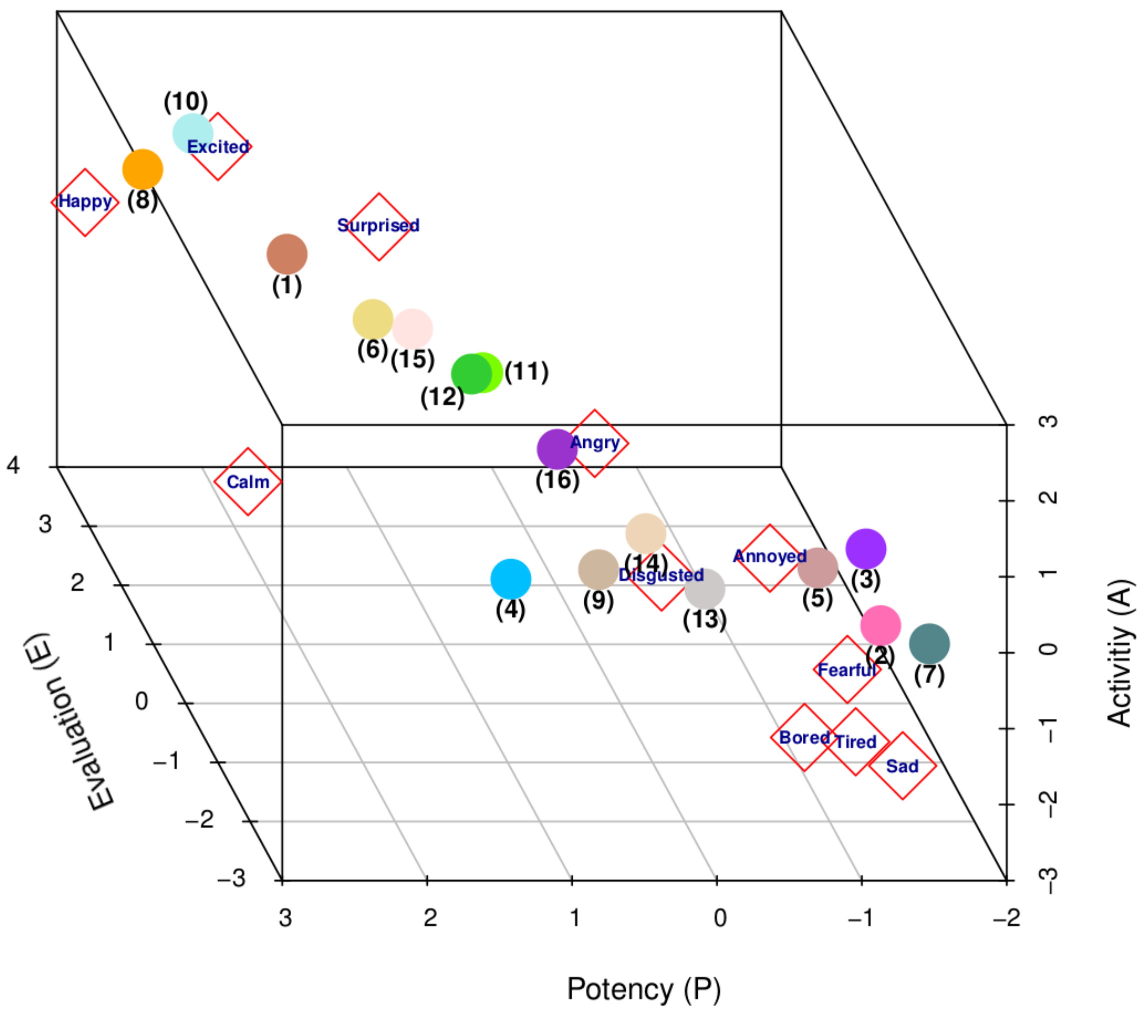


Figure 4.2: Visualization of participants' average EPA ratings for each sentence (shown with dots) and EPA ratings of the 11 emotions taken from the EPA database (shown with rhombi). Numbers show the situations (see Table 4.1 for the associated sentences). For example, EPA values for the sentence with number 10 (I think I heard someone is calling for help, we might have found a survivor) is closest to the EPA value of "excited" emotion.

Table 4.4: Predicted emotions based on the distance between average EPA ratings of participants' EPA scores and EPA ratings of the 11 emotions from the U.S.A 2015 dataset [191]. Corr shows the Pearson's correlation coefficient [16] between EPA ratings of participants from Canada and USA. Dist. shows the calculated euclidean distance [42] between EPA values of the predicted emotions and participants' ratings. The last column contains the mappings obtained in Experiment 1 for comparison.

Sentences	**Average**			**Corr**	**1^{st} Prediction**		**2^{nd} Prediction**		**From Exp. 1**
	E	**P**	**A**		**Dist.**	**Emotion**	**Dist.**	**Emotion**	
I can again communicate with our team outside of the building	2.55	1.73	0.93	0.98	1.39	Excited	1.49	Happy	excited, **happy**, **calm**
I lost communication with our team outside of the building so we are on our own now.	-2.00	-1.35	-0.42	0.99	0.56	Fearful	1.23	Annoyed	**fearful**, **annoyed**
I am stuck and might need help to proceed	-1.08	-1.46	-0.13	0.96	1.48	Fearful	1.49	Annoyed	fearful, **annoyed**
I detected dangerous material here, let's proceed carefully	-1.03	0.98	-0.56	0.88	1.76	Disgusted	2.17, 2.43	Annoyed, Fearful	**fearful**
I believe we are behind schedule. I also noticed it is getting dark and there is not much time left	-1.72	-0.98	0.12	0.99	0.68	Annoyed	1.06	Fearful	**fearful**, **annoyed**
I found an item that could belong to a person. Maybe the person is nearby	2.16	1.23	0.38	0.99	1.94	Surprised	2.16	Excited	happy, **excited**, **calm**
My battery is running low and I will stop working soon	-1.80	-1.73	-0.82	0.98	0.90	Fearful	1.35	Sad	sad, **tired**, **fearful**
I think I found a surviving person	3.00	2.63	1.70	0.96	0.77	Excited	0.94	Happy	surprised, **excited**, **happy**
I detected that there might be a risk of further collapse so we should only proceed with caution	-0.99	0.37	-0.47	0.42	1.57	Disgusted	1.75, 1.99	Annoyed, Fearful	**fearful**
I think I heard someone is calling for help, we might have found a survivor	2.81	2.32	2.32	0.98	0.20	Excited	1.65	Happy	surprised, **excited**, **happy**
I think we need additional team members	0.83	0.77	0.71	-0.29	1.68	Surprised	2.79	Excited	NA
I think we have more team members than we need. One of us should join the other team	0.65	0.88	0.83	0.22	1.61	Surprised	2.62	Angry	NA
I detected that the temperature of the environment is too cold for a person	-1.25	-0.30	-0.52	0.99	1.35	Fearful	1.36	Annoyed	NA
I detected that the temperature of the environment is too hot for a person	-1.45	0.14	0.36	0.99	0.83	Disgusted	0.97	Annoyed	NA
I found a victim that requires medical attention	0.06	1.42	1.88	0.97	1.39	Surprised	2.02	Angry	NA
I detected that this rescue route requires obstacle clearance	-0.25	0.49	0.54	0.94	1.98	Angry	2.04	Disgusted	NA

We also further analyzed the correlations between EPA ratings of participants from Canada and the USA. While the ratings were generally highly correlated for most of the situations, we observed that for three of the situations correlation value was lower (-0.29, 0.22, and 0.42). These sentences with lower correlation are shown in blue in Table 4.4. To emphasize the differences, mean EPA ratings of these sentences for each country are presented in Table 4.5.

Table 4.5: The sentences which have either low or no correlation between participants from Canada and U.S.A.

Sentence	Location	(E)	(P)	(A)
I detected that there might be a risk of further collapse so we should only proceed with caution	Canada	-0.83	0.46	-1.04
	U.S.A.	-1.18	0.26	0.20
I think we need additional team members	Canada	0.58	0.66	0.54
	U.S.A.	1.12	0.89	0.92
I think we have more team members than we need. One of us should join the other team	Canada	0.48	0.61	0.74
	U.S.A.	0.85	1.21	0.94

4.3.4 Discussion

We argue that emotions can provide SAR robots with a communication channel that can complement the existing communication modalities and improve the success and efficiency of communication. Further, such social ability might help rescue workers to be less affected by Post Traumatic Stress Syndrome (PTSD) by reducing the frequency or intensity of perceived stress levels of dangerous situations encountered during missions [183]. Victims can also benefit from those emotions since social SAR robots can contribute to a reduction of stress levels of victims and prevent shock as suggested in [20]. After confirming the feasibility of using emotions in SAR robots via Experiment 1, we conducted Experiment 2.

To address **RQ3** and check whether the obtained mappings via Study 1 (Chapter 3) might be affected by the selected set of emotions, we asked if it is possible to use a method to quantify the emotions and get the ratings in a way that it is independent of the selected set of emotions. Therefore, we used ACT and asked participants to rate the sentences on three different dimensions: Evaluation, Activity, and Potency (EPA) in this study. We then used the EPA values associated to the set of 11 emotions in Experiment 1 to check whether the mappings would be consistent with mappings in Experiment 2. EPA values for the 11 emotions were taken from the U.S.A 2015 dataset [191], which is the most recent dataset on EPA ratings. We also added more sentences in Experiment 2 to cover situations relevant to different types of SAR operations and robot emotions.

Results showed that the emotions obtained through Experiment 1 and Experiment 2 were consistent (see Table 4.4), suggesting that using EPA ratings for the sentences can lead to similar results. Also, other than not being dependent on a specific set of emotions, this method has the potential to be automated in the future. Also, consistency in results demonstrated that the obtained mappings were independent of whether emotions were provided as a direct choice to participants (as in Experiment 1) or calculated using ACT. It is important to stress the findings related to the two sentences "I detected dangerous material here, let's proceed carefully" and "I detected that there might be a risk of further collapse, so we should only proceed with caution". For these sentences, the third closest emotion resulted in consistent mappings as in the first experiment. The reason behind this might be related to the first closest emotion found disgusted, which was not suggested by the mapping obtained from the first experiment at all. Nonetheless, it is not possible to make further comments about the findings of these two experiments with the current data.

Employing a dimensional emotion model like in Experiment 2 is not common with sentences. These values usually exist for emotions, identities, and actions. Using these dimensions with sentences has complications. For example, the mappings will be highly context-dependent. It would be hard (if not impossible) to conduct extensive surveys to gather ratings for all combinations of sentences, similarly that the other EPA values are collected (as a large number of sentences can be created with the combination of the related words). Further, there is currently limited literature on mapping sentences to dimensional emotion models, most of which can only evaluate sentences on the Evaluation dimension (i.e., a sentence's sentiment). However, using such models can be promising in the future and lead to mappings that are independent of a limited set. Furthermore, the set of emotions can be changed depending on what emotions a social robot in SAR is capable of showing, and the same ratings can be used for obtaining the new mappings.

It is important to emphasize that, while the EPA ratings of the sentences are independent of a set, we still need to decide on a set of reasonable emotions for the mappings. The ACT datasets include a large set of emotions, which may not be relevant in any specific context. For example, if we had used the complete set instead of limitings it to our set of 11 emotions (i.e., to use the "Emotion" dataset of U.S.A 2015), the two closest affective expressions for the sentence "I detected dangerous material here, let's proceed carefully" would have been "obligated" and "aggrieved", which may (a) not be appropriate for the context of SAR, and (b) not be feasible to express on a social robot.

Furthermore, for this study, we presented sentences that a robot would say during SAR situations to the participants as text rather than using a specific robot that says these sentences. The reason behind this choice was to avoid the possible effect of a particular robot's embodiment, voice, etc., on participants' responses. This approach allowed us to investigate the feasibility of matching emotions with SAR related situations without biasing participants to any specific robot/platform since the focus of these two studies is not on implementing these affective expressions on SAR robots, but rather investigating the possibility of using emotions to convey specific messages in SAR context.

Finally, we noticed that in three of the situations, the ratings obtained from the US and Canada were either not correlated or only had a low correlation (See Table 4.5). At the same time, it was not the case for the rest of the situations. One possibility is that the cultural differences were reflected in these situations more than the others. For example, participants in the US rated the sentence “I think we need additional team members” as *better*, more powerful, and more active, as compared to the participants in Canada, who rated this sentence closer to neutral in all dimensions. It was very similar to the sentence “I think we have more team members than we need. One of us should join the other team”, and in both cases, it involved a change in the structure of teams. Future work is needed to study why the ratings were not correlated for these few situations and if cultural differences were in fact the reason for this observation. If this is the case, it may suggest that cultural differences should also be taken into account while designing emotions for communication of SAR robots.

As these studies confirmed the feasibility of bringing emotions into the search and rescue context, we propose using emotions as an additional communication channel to the existing ones such as video streams, voice, and text [92]. The intention behind proposing an additional interaction modality is to complement existing multi-modal channels rather than replacing them. In this way, we would be able to employ SAR robots that have more robust and failure-safe communication abilities that might help to improve field workers' shared mental model and situational awareness [185]. Experiment 2 suggested an approach that can help with obtaining these mappings more efficiently.

4.4 Limitations and Future Work

This study had several limitations. First of all, due to the online nature of the studies, participants did not have a chance to interact with real SAR robots. They also did not experience a SAR scenario, which could help with understanding the situations and affect the mappings. While illustrating possible SAR operations with several pictures of

SAR robots, the obtained results might differ in real-life scenarios and if the mappings were obtained from participants who had experience with SAR. However, such an online approach was followed as the first step for this direction of research in order to not bias participants with the appearance of a particular robot, and it also helped with reducing the experimenter bias [70]. Furthermore, this approach has been shown to be effective in many HCI and HRI studies and has gained more attention since COVID-19 has affected the feasibility of conducting in-person HRI studies. Nonetheless, future work is needed to investigate if/how obtained mappings would translate to real-life situations and if rated by participants who have experience with SAR situations.

Secondly, although participation was limited to North America, participants' level of English was not tested during the study. Yet, based on their answers to attention check questions, it is reasonable to assume that they understood the task and the sentences.

While this study (different from the first one) led to ratings that can be used with varying sets of emotion, we did not examine how the mappings changed based on different emotion sets. Future work could obtain mappings using different emotion subsets, for example, those that a specific robot can show.

After making sure that the obtained mappings are robust and ready to be used for real-life scenarios, implementing the emotions on robots that have been already used in SAR operations would be a natural next step for future work. For this purpose, following existing design guidelines for SAR robots suggested in [20] might be helpful. Guidelines for implementing such expressions on non-zoomorphic (compared to e.g. [66]) can also help with the implementation of these emotions.

Lastly, further studies that employ emotions to convey information from robots to humans in different application areas, such as firefighting and service robotics, can help support the generalizability of using emotions as a communication channel to complement multi-modal human-robot interaction in other similar contexts. This would help to better investigate context dependency of perceived emotions.

4.5 Conclusion

- The goal for this experiment was to find mappings between SAR related sentences and emotions in a way that it is independent of the emotion set.
- Different from the first experiment, participants in this experiment were asked to rate EPA dimensions suggested by ACT for SAR related sentences instead of selecting emotions through a list.

- Obtained mapping between sentences and emotions was consistent with the one obtained in the first experiment.

- This approach allowed us to obtain mappings independent of the emotion list which might be needed if the findings are used for different robots that are capable of expressing different emotions or different fields that require different emotion sets depending on the context.

Chapter 5

Study 3: Implementing Affective Expressions on a Appearance-Constrained Robot and Improving Human-Robot Interaction in Search and Rescue Teams

5.1 Research Questions and Hypotheses

This study aims to address the following research questions.

RQ4 How can affective expressions be designed and implemented on appearance constrained SAR robots using lights?

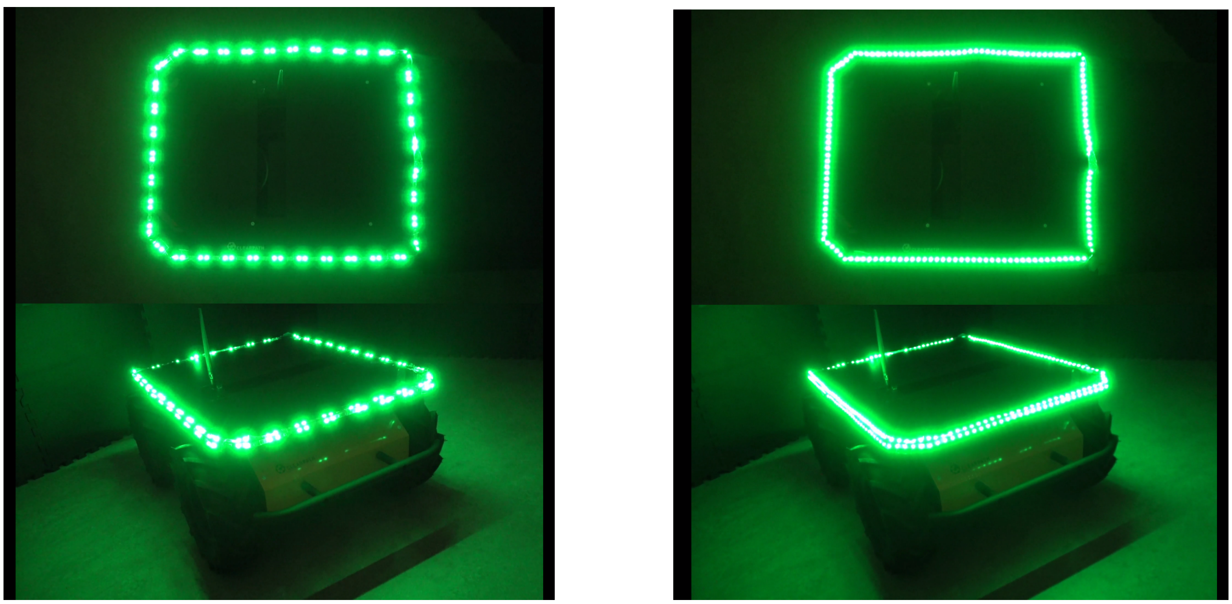
RQ5 Can affective expressions complement and improve multi-modal communication in human-robot SAR teams?

Based on the discussion in the sections 2.2 and 2.1.2, our hypotheses are:

H1 Affective expressions will increase participants situational awareness (i.e. their perception of what is happening in the disaster area) when other communication modalities like text fail.

H2 Affective expressions will decrease participants response time (i.e. how fast they perceive what is happening or how fast they take an action) by complementing other communication modalities.

5.2 Methodology



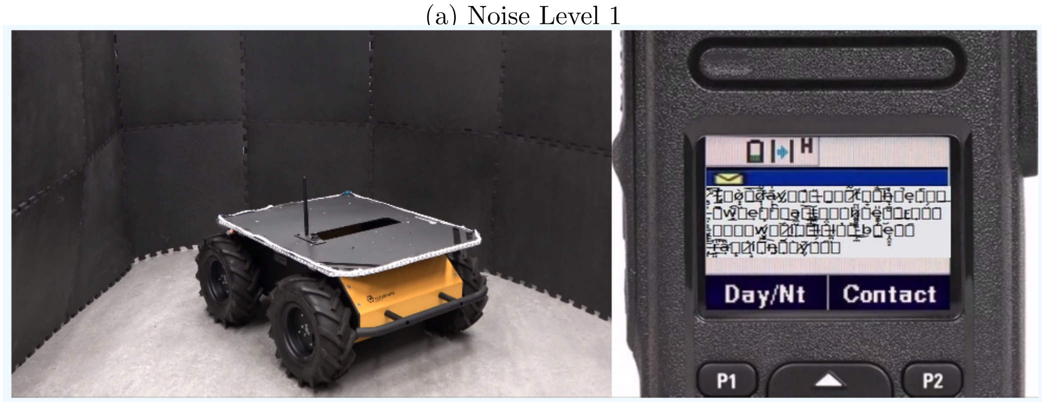
(a) Two LED Pattern

(b) Full LED Pattern

Figure 5.1: Screenshots from a pilot study that shows different light patterns used to express"happy". They were designed based on EPA dimensions [78], where Evaluation (E) dimension represents the color of the light, Potency (P) dimension represents the intensity of the light, and Activity (A) dimension represents the frequency of the light.

In order to address **RQ4**, affective expressions were designed and implemented on Clearpath's Husky robot using LED strips and based on EPA dimensions suggested by ACT [78]. Informal pilot studies were conducted with the recorded videos of Husky's affective expressions with lab members. In each pilot, different parameters of these expressions (light intensity, frequency, color, lighting pattern, EPA scale mapping, etc.) were varied systematically. One can see the affective light designs from an earlier pilot study in Figure 5.1. The reason behind the pilots was to understand how to better implement natural and perceivable affective expressions on Husky using lights. After analyzing the results of the pilot studies, we decided to continue with a full light pattern (i.e. turning on/off all the

(a) Noise Level 1



(b) Noise Level 4

Figure 5.2: Glitchy versions of the text message "Today the weather will be rainy" displayed on radio transmitter device Motorola XPR 7550e and Clearpath's Husky robot. Noisy messages were created thanks to Zalgo text generator [31].

LEDs on the strip at the same time) since pilot data indicated better perception accuracy from all viewing angles. In addition, we determined the color of the lights should represent the goodness of the message (Evaluation) while the brightness of the lights should represent the power of the message (Potency) and the frequency of the lights should represent the activity of the message (Activity) following ACT theory. The details about the final implementation of expressive lights were explained in Section 5.3.1.

After implementation of affective lights, noisy text messages were created in order to mimic the situation where the communication modality fails during the SAR mission. Zalgo text with different chaos levels was used to make the text messages glitchy [31]. Another pilot study with the members of the Social and Intelligent Robotics Research Laboratory was conducted to decide the noise level. Participants were presented to different noisy versions of the text "Today the weather will be rainy", and they were asked to guess what was written in the text. In order to create a realistic scenario, the noisy text messages were displayed on a radio transmitter device (Motorola XPR 7550e), which is widely used during real SAR missions [151]. Figure 5.2 shows some versions of the noisy text message shown to participants during the pilot. After analysis of the pilot results, noise (i.e. chaos) level 3 was selected as the minimum level which makes the text unreadable.

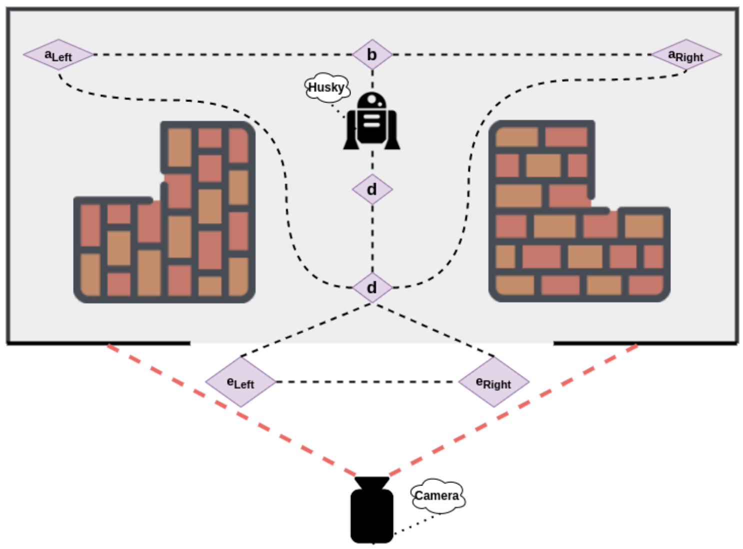
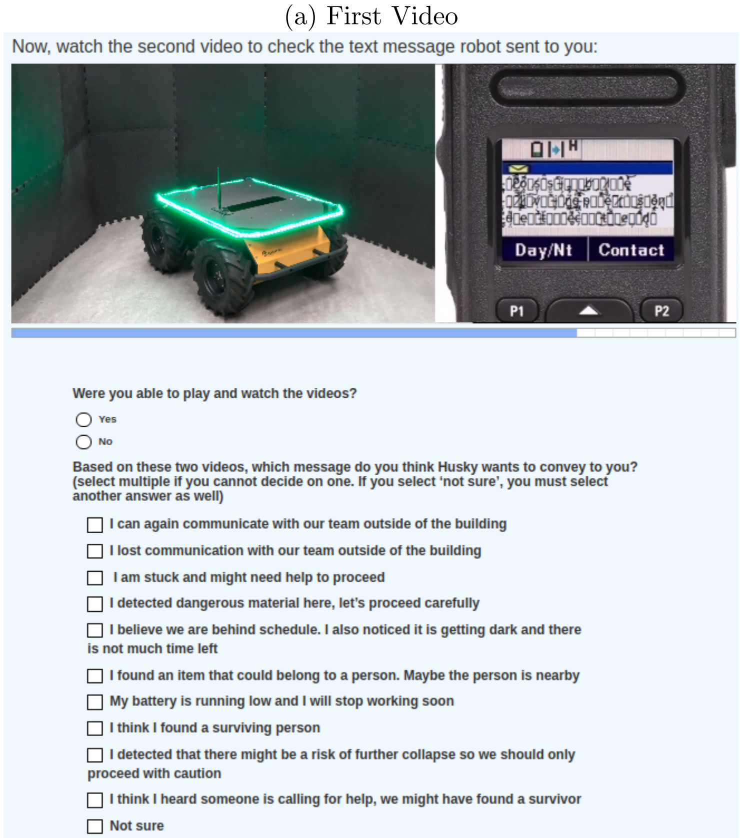


Figure 5.3: Experimental Setup used to mimic common scenarios happening during SAR missions in Experiment 3. Different locations were shown with pink rhombuses and labeled with letters. Possible routes between these points were shown with dashed lines. For each scenario used in the experiment, Husky starts its movement from one of these points and visits particular points using dashed routes. These movements were recorded by the camera and shown to participants as videos.

The main experiment had a between-subject design where participants were divided into two conditions: emotion condition and no emotion condition. While Husky expressed emotions using lights in emotion condition, it did not use any lights in no emotion condition (i.e. lights were turned off all the time). After an initial training step to teach affective expressions of Husky, participants were shown 2 videos (the first one shows movements of Husky in the simulation and the second one shows the noisy message with/without emotional lights) for each scenario (10 x 2 = 20 in total) and they were asked to guess the message Husky wants to convey in Figure 5.4. The mapping between these messages and emotions was obtained from the first study (see Section 3)

The experimental setup shown in Figure 5.3 was used in the experiment in order to simulate common scenarios happening during SAR operations. In this setup, Husky starts from a particular point depending on the scenario simulated and follows a specific route. For example, for the scenario "I think I found a surviving person", Husky starts from

(a) First Video



(b) Second Video and Questions (Emotion condition since Husky's lights are active)

Figure 5.4: The main task in Experiment 3 where participants were shown two videos for each scenario and then asked to guess what Husky wants to convey considering both videos

Table 5.1: Different routes were followed by Husky for each scenario. Please see Figure 5.3 for experimental setup and locations of labelled points ($a_{Right}, a_{Left}, b, c, d, e_{Right}, e_{Left}$). Matched emotions for each scenario are given under the Emotion column.

No	**Scenario**	**Path Followed**	**Emotion**
1	I can again communicate with our team outside of the building	$a_{Right} - b - c - d$	happy
2	I lost communication with our team outside of the building	$b - d - e_{Right} - e_{Left} - d$	fear
3	I am stuck and might need help to proceed	$b - c - d - c - d - c - d$	annoyed
4	I detected dangerous material here, let's proceed carefully	$b - a_{Left} - b - c - d$	fear
5	I believe we are behind schedule. I also noticed it is getting dark and there is not much time left	$b - c - d$	annoyed
6	I found an item that could belong to a person. Maybe the person is nearby	$b - a_{Right} - b - c - d$	excited
7	My battery is running low and I will stop working soon	$a_{Left} - b - c - d$ (decrease speed gradually)	tired
8	I think I found a surviving person	$b - a_{Right} - d$ (use curvy path from a to d)	excited
9	I detected that there might be a risk of further collapse so we should only proceed with caution	$b - a_{Left} - d$ (use curvy path from a to d)	fear
10	I think I heard someone is calling for help, we might have found a survivor	$a_{Right} - d$ (use curvy path from a to d)	excited

point b then goes to point a_{Right}. During this movement, Husky slowly goes out of the point of view. Then, it appears again and moves toward point d by taking the curvy path (shorter path comparing to going first point b and then point d). During these movements, participants were also notified that they got a text message from Husky (without showing the message itself) and the time they get this text notification was controlled and differs in each scenario. For example, for the scenario "I think I found a surviving person", participants got notified with beeps once Husky is in point a_{Right} and out of the point of view so that participants will be informed that something happened while Husky was around that point. This provides additional context information to participants about the scenario itself. Different paths followed by Husky for each scenario are shown in Table 5.1. Simulation videos for all scenarios can be watched through this link.[1]

[1] https://www.youtube.com/playlist?list=PL2L3Ul8a1xkwRlxJuSPic78RtCtDYg01E

5.3 Experiment 3

5.3.1 Robotic Implementation

Design of Emotional Lights

A NeoPixel RGB LED strips were used to design affective expressions. All emotions that were used in the experiment were programmed on Arduino micro-controller in C++ using the associated LED strip libraries: Adafruit NeoPixel[2] and FastLED[3]. Each emotion was given a function in which the period, wavelength, and color of the wave could be altered based on the design of the emotion. For the experiment, EPA dimensions in ACT was transformed in order to represent different parameters of LED lights (see Table 5.2 for description of parameters and Table 5.3 for parameter values). Related software was made open-source in order to provide an example for researchers who are interested in implementing affective expressions based on ACT.[4]

Table 5.2: Description of parameters to transform EPA ratings into corresponding LED attributes. We determined the range of LED parameters based on the feedback through previous pilot studies.

	EPA Values			LED Parameters		
Name	**Description**	**Min**	**Max**	**Description**	**Min**	**Max**
Evaluation (E)	How good (Bad vs Good)	-4.3	4.3	Color	Red	Green
Potency (P)	How powerful (Weak vs Powerful)	-4.3	4.3	Intensity	0	255
Activity (A)	How active (Passive vs Active)	-4.3	4.3	Duration	4300 msec	300 msec

For the experiment, we attached two LED strips on the Husky robot's top and side (360°) to give better perception from various angles like in Figure 5.5. Afterward, videos of the robot expressing different emotions were recorded to be used during the experiment.[5]

Search and Rescue Simulation

Gazebo simulator was used with Robot Operating System (ROS) middleware to create a realistic search and rescue simulation [169]. ROS packages provided by Clearpath Robotics was employed to simulate Husky in Gazebo.[6]. To construct the SAR disaster environment, various 3D models provided by Open Robotics[7] were combined based on experimental

[2] https://github.com/adafruit/Adafruit_NeoPixel
[3] https://github.com/FastLED/FastLED
[4] www.github.com/samialperen/epaLights
[5] https://youtube.com/playlist?list=PL2L3Ul8a1xkwvAhQ4jBc6OOwqJTGSifHU
[6] https://github.com/husky/husky
[7] https://github.com/osrf/gazebo_models

procedure. Resulted simulation environment in Gazebo was made publicly available.[8] Simulated disaster environment and virtual Husky robot can be seen in Figure 5.6.

Table 5.3: Values of parameters to transform EPA ratings into corresponding LED attributes for each emotion used in the study. EPA values for the emotions were obtained from the USA Student 2015 dictionary [191]. We normalized the range of LED parameter values based on the EPA range of emotions considered. For example, emotion "happy" has the largest P value (2.85), so it was converted to 255 (max LED value for RGB parameters). We took the risk of a seizure into account while selecting the minimum value of LED lights duration [215].

	EPA Values [191]			**LED Parameter Values**			
Emotion	**E**	**P**	**A**	**R**	**G**	**B**	**Duration (msec)**
tired	-1.58	-1.28	-2.28	31	0	0	4154
happy	3.54	2.85	0.85	0	255	0	1609
fear	-2.41	-1.07	-0.81	54	0	0	2958
excited	2.77	2.13	2.46	0	174	0	300
annoyed	-2.13	-0.47	0.58	64	0	0	1828

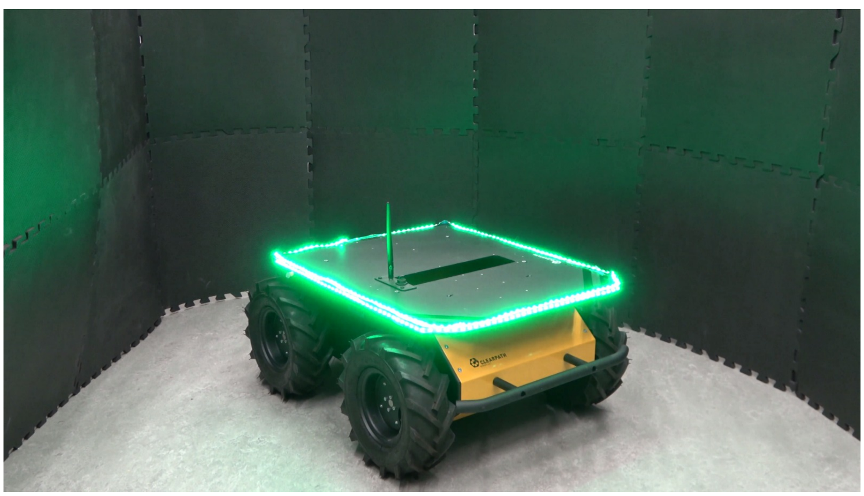


Figure 5.5: Clearpath Husky robot with attached LED strips

5.3.2 Procedures and Measures

The experiment has three steps as summarized in Figure 5.7. All participants in each condition (emotion vs. no emotion) followed the same procedure except the second part of the main task. The overall interface used in the experiment can be seen in Appendix D.

[8] https://github.com/samialperen/husky_sar

Step 1 - Training: Participants who signed the consent form moved to the training step. They were instructed to watch the training video carefully and as many times as needed to learn the emotional displays of the Husky. In the training video, Husky expressed emotions with the help of lights, and the name of each emotion was clearly shown at the bottom left corner of the video. Five different emotions (happy, excited, tired, annoyed, fear) were implemented in this study. This emotion set was decided based on the mapping obtained in the first study (see section 3).

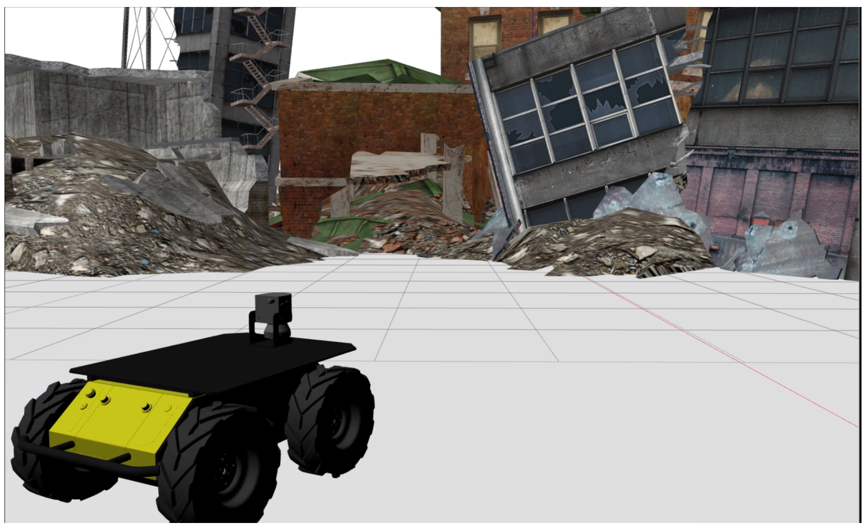


Figure 5.6: The simulated Search and Rescue Environment used in the study with the Clearpath Husky Robot

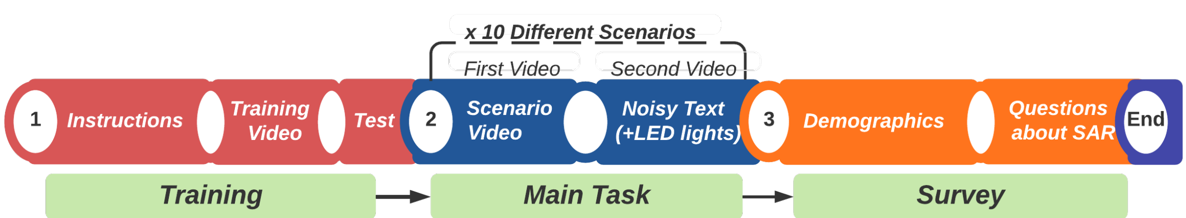


Figure 5.7: Summary of Experimental Procedure Followed in the Experiment 3

Upon watching the training video, separate videos for each affective expression were shown in random order without displaying the name of the expression, and participants were asked to guess what expression was shown. If participants predicted the emotion correctly, that emotion was marked as passed, and another emotion was shown next until they predict all the emotions accurately. Otherwise, the emotion they failed to recognize was shown again randomly during the training. Participants who saw more than “10” test videos (i.e., seeing each emotion two times on average since there are five different emotions in total) were considered as failed in the training step. They were still allowed to continue the rest of the study, but their data was labeled as “failed” to track their performance during the analysis.

This training step is a crucial part of the experiment since it allows us to have a greater confidence that participants can perceive the affective expressions of Husky correctly. As stated in the previous sections, the focus of this study is to investigate whether affective expressions can help to increase situational awareness of search and rescue workers, not to evaluate recognition of emotions. Although we were hoping that to some extent the affective expressions that we designed were intuitive to understand, in practice SAR workers get training regularly, including how to use new tools, so adding this step to their regular training routine might actually be feasible in future applications.

Step 2 - Main Task: Participants who either completed the training successfully (labeled as “passed”) or failed and hit the threshold (labeled as ”failed”) moved to the instructions page for the main task. They were told that they are part of a robot-assisted search and rescue team, and they are called to duty. After watching an introduction video to search and rescue scene, they moved to the next instruction page, where the procedure for the main task was explained. They were told that they will see two videos for each scenario: (a) the first video is in the simulation environment and called scenario video and (b) the second video is in real-life that shows noisy text message (plus affective lights of Husky if in emotion condition). They were given a complete list of possible messages, and they were told that their task is to choose what message Husky wanted to convey to them.

In total, each participant watched 20 videos for 10 different SAR scenarios (see Table 5.1 for all the scenarios). These videos were shown in a random order, and participants had an option to select multiple messages in the given list to answer if they are not sure. They had an option to replay all the videos. The videos were paused if they switched the interface tab or opened another application. In this way, we measured their response time accurately.

Step 3 - Questionnaire: After completing the main task, participants were moved to the survey step, where they first answered questions about demographics and then about SAR (see Appendix D for the whole interface).

For the demographics questionnaire, participants answered questions about their gender, age, education, and ethnicity. They had the freedom not to answer any of the questions.

Between demographics and SAR-related questionnaires, participants answered one of the attention check questions. Then, they responded to questions related to the design of the experiment. First, they were asked to rate the difficulty level of noisy text messages they saw during the study. Second, they were given a picture with another noisy text message, and they were asked to write down the message. Thanks to these two questions, we evaluated the difficulty of the noise level used in the study.

Finally, participants were asked to report their opinion about (1) how useful they think rescue robots are, (2) how familiar they are with rescue robots, (3) whether they had seen a SAR robot before, (4) how necessary they think rescue robots are, (5) how much they believe rescue robots will be better than rescue dogs in the future, (6) how not useful they think rescue robots are, (7) how much they believe we won't need rescue robots in the future and (8) another attention check question. After these questions, they answered another one depending on the condition they were assigned. Participants in the emotion condition reported (9) how much they think the robot's use of lights in order to convey emotions was helpful to guess messages sent by the robot. On the other hand, participants in no emotion condition stated (9) how much they would prefer the robot to use lights in order to convey emotions which might be helpful to guess messages sent by the robot. All questions in this section were on a continuous scale, and participants had an option "prefer not to share" if they did not want to answer any of the questions.

5.3.3 Participants

We recruited a total of 151 participants for the study on the Amazon Mechanical Turk platform. In order to increase the quality of obtained data, only participants whose approval rate is higher than 97% based on at least 100 HITS allowed to join the study. Recruited participants were located either in the USA or Canada. Participants who completed the study were paid $3 for compensation, while a pro-rated amount was paid for those who did not finish the task. This study received full ethics clearance from the University of Waterloo's Research Ethics Committee (see Appendix A).

Recruited participants were filtered based on attention check questions. Sixteen of them failed in the attention check question "I think drinking water is liquid". Moreover, data related to 33 participants who gave inconsistent responses (i.e., not selecting opposite sides) to the questions "I think rescue robots are useful" and "I think rescue robots are not useful" discarded. After filtering, 102 participants (37 female, 65 male; ages 22-69, avg: 38.9, std: 11.1) left for the analysis where 53 of them were in emotion condition, and 49 of them were in no emotion condition.

Table 5.4: List of considered factors in Linear Mixed-effects Model (LMM) analysis as well as survey questions presented at the end of the experiment (factors marked with ^{x} in the table)

Factor	**Description**	**Data Type**
Gender ^{x}	Gender of participants	Discrete
Age ^{x}	Age of participants	Number
Education ^{x}	Highest level of education of participants	Discrete
robot_useful ^{x}	I think rescue robots are useful	Scale 0–1000: Completely Disagree / Agree
familiar_sar ^{x}	I was familiar with rescue robots before this study	Scale 0–1000: Completely Disagree / Agree
seen_sar_robot ^{x}	I had seen an example of a rescue robot before this study	Scale 0–1000: Completely Disagree / Agree
robot_necessary ^{x}	I think rescue robots are necessary	Scale 0–1000: Completely Disagree / Agree
better_than_dogs ^{x}	I believe in the future rescue robots become better than rescue dogs	Scale 0–1000: Completely Disagree / Agree
robot_not_useful ^{x}	I think rescue robots are not useful	Scale 0–1000: Completely Disagree / Agree
robot_no_need ^{x}	I believe in the future we will not need rescue robots	Scale 0–1000: Completely Disagree / Agree
Attention Check ^{x}	I think the drinking water is liquid	Scale 0–1000: Completely Disagree / Agree
helpful_lights ^{x} (emotion cond)	I think the robot's use of lights in order to convey emotions was helpful to guess messages sent by the robot	Scale 0–1000: Completely Disagree / Agree
prefer_lights ^{x} (no emotion cond)	I would prefer the robot to use lights in order to convey emotions which might be helpful to guess messages sent by the robot.	Scale 0–1000: Completely Disagree / Agree
hard_text ^{x}	How hard was it to read the noisy text messages?	Scale 0–1000: Completely Disagree / Agree
Perception Accuracy	Shows whether participants guessed the particular scenario correctly	Boolean
Perception Time	Total active time in seconds that participants spent for each scenario (excluding inactive time)	Number (seconds)
Inactive Time	Total inactive time that participants did not pay attention to the task (task was on pause during these times)	Number (seconds)
switchTabCounter	Total number of tab switches occurred during the task	Number
Condition	Condition that participants were assigned to	Discrete: emotion vs no emotion

5.3.4 Statistical Analysis

In this experiment, we investigated two main metrics: perception accuracy and response time. Perception accuracy of participants was calculated by measuring their success in guessing SAR related messages. Response time was reported by measuring how fast they predicted the messages. On the other hand, independent measures considered in this study are: (a) participants' response to the questions in the survey, (b) the order of messages seen by the participants, (c) the total number of times they switched from the main task, (d) the total inactive time that they did not spend on the task and (e) the condition that participants belong to (emotion vs. no emotion). List of all considered factors was given in Table 5.4. To investigate the relation between the independent and dependent measures, LMM [13] was employed based on minimizing Akaike's Information Criterion (AIC) [9]. Moreover, one-way binomial tests were applied assuming uniform probability distribution as the null hypothesis to determine whether participants selected a specific scenario (or emotion in training step) significantly over another one [178].

5.3.5 Results

During LMM analysis, it was detected that participants in emotion condition had significantly more perception accuracy than participants in no emotion condition ($se = 0.04, t = 2.287, p = .024$). On the other hand, no significant correlation was found between the condition participants assigned to and response time ($se = 6.06, t = -0.05, p = .96$). One can check full LMM results in Table 5.5 and Table 5.6.

Table 5.5: Linear Mixed-effects prediction model with minimum AIC to measure participants' perception accuracy.

Covariate	**Perception Accuracy**			
	Estimate	SE	t	Pr ($> \lvert t \rvert$)
Condition				
No Emotion [b]				
Emotion	0.09	0.04	2.287	0.024 *
Familiar with SAR	-1.64e-04	6.78e-05	-2.417	0.017 *
SAR Robots Not Useful	-2.59e-04	8.02e-05	-3.234	0.002 **

* = $p < .05$; ** = $p < .01$; *** = $p < .001$; [b] = baseline level

LMM analysis resulted in interesting findings. As suggested in Table 5.5, participants who thought rescue robots are not useful or who were familiar with SAR had a lower per-

Table 5.6: Linear Mixed-effects prediction model with minimum AIC to measure participants' response time.

Covariate	Response Time			
	Estimate	SE	t	Pr $(>\|t\|)$
Condition				
No Emotion [b]				
Emotion	-0.3	6.06	-0.05	0.96
Inactive Time	0.11	0.03	3.273	0.001 **
SAR Robots Useful	0.04	0.02	2.012	0.047 *
Familiar with SAR	0.06	0.01	4.355	<0.001 ***
Seen SAR Robot Before	-0.04	0.01	-2.865	0.005 **
SAR Robots Necessary	-0.03	0.02	-1.543	0.126
Order Number	-6.1	2.45	-2.487	0.013 *

* = $p < .05$; ** = $p < .01$; *** = $p < .001$; [b] = baseline level

ception accuracy. Moreover, Table 5.6 gives thought-provoking results. Participants who were familiar with SAR or participants who think rescue robots are useful or participants who had a higher inactive time spent more time to predict the shown scenarios. On the other hand, participants who saw a rescue robot before, who think rescue robots are necessary were faster to respond. We also detected that participants tended to get faster to respond shown videos as they see more videos (ordering effect).

For the recognition of emotions itself during the training step (i.e., training success), 41 participants (22 in emotion condition and 19 in no emotion condition) out of 102 (53 in emotion condition and 49 in no emotion condition) could not learn all the emotions and hit the threshold limit of training. They were allowed to move into the next step of the study without learning all the emotions. They are labeled as "failed", and their data was investigated separately depending on the analysis.

During the rest of this section, findings regarding training success & design of affective lights, perception accuracy, and response time for the main task were shared in more depth.

Training Success

Training success is related to participants' overall accuracy in perceiving Husky's emotions after watching the initial introductory training video. Figure 5.8 shows the perception accuracy for each emotion. All emotions were perceived with accuracy more than the chance

(20%). While "happy" was perceived with the best accuracy ($\approx$ 91%), and "annoyed" with the worst accuracy ($\approx$ 59%).

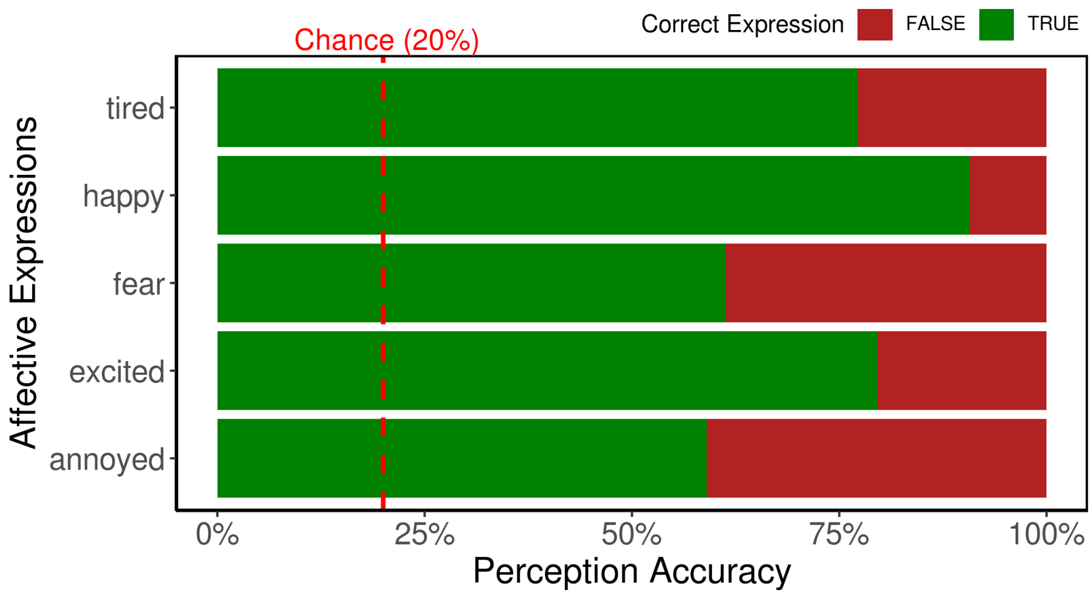


Figure 5.8: The success of participants to predict each emotion during the training step

Once participants mispredict an emotion, that emotion is shown randomly later until they succeed. However, they can not see more than ten emotion videos (for five different emotions). Once they hit this threshold, they were allowed to move into the next step. Nonetheless, they were labeled as "failed" in training.

Table 5.7: The table shows participants' incorrect guesses to recognize implemented affective expressions on Husky. Rows show emotions expressed by Husky, and columns show participants' corresponding responses to these expressions. Since only incorrect responses of participants who failed to learn all emotions in the training step included (41 participants failed in total; 22 in emotion condition, 19 in no emotion condition), all diagonal entries have zero, i.e. having a non-zero diagonal entry means that emotion perceived correctly.

	Participants' Responses	**Tired**	**Fear**	**Excited**	**Happy**	**Annoyed**
Emotions	Tired	0	7	4	4	10
	Fear	26	0	4	4	14
	Excited	0	1	0	20	1
	Happy	2	1	5	0	1
	Annoyed	14	27	7	4	0

In this experiment, 41 participants failed to learn all the emotions and hit the threshold. Table 5.7 shows only incorrect responses of these 41 participants. Among the misunderstood emotion pairs, fear-tired and annoyed-fear are the ones that got confused most, while the happy-excited pair is the least confused one.

Emotions that caused participants to fail in the training and how many attempts each participant guessed them incorrectly were shown in Figure 5.9. Fear and annoyed emotions have the largest average number of attempts in both conditions (emotion vs. no emotion).

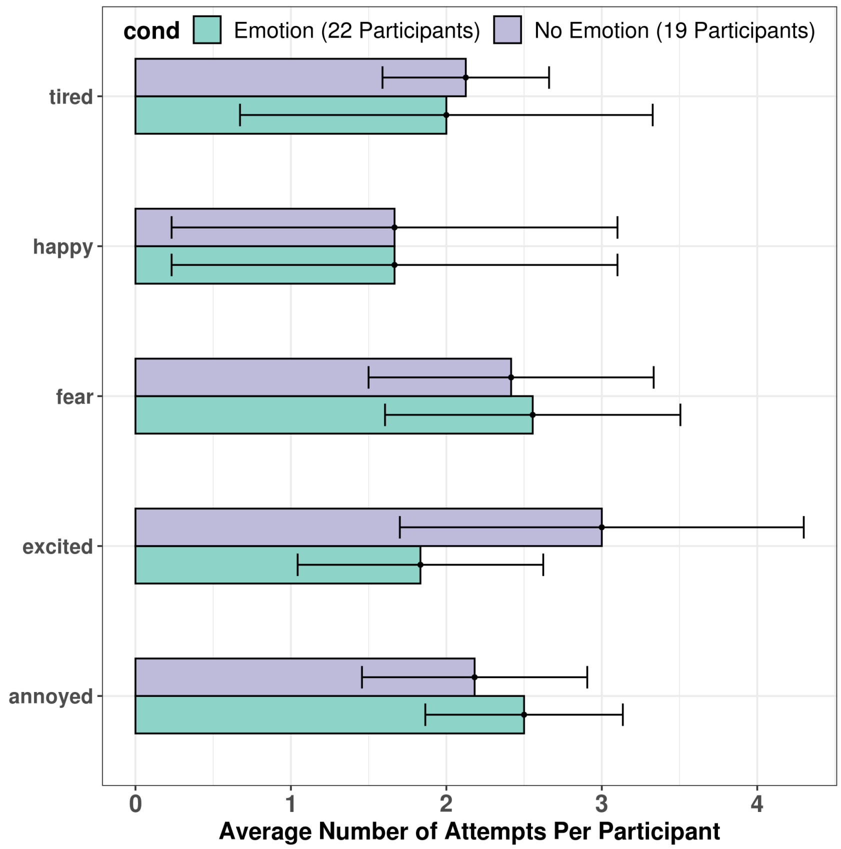
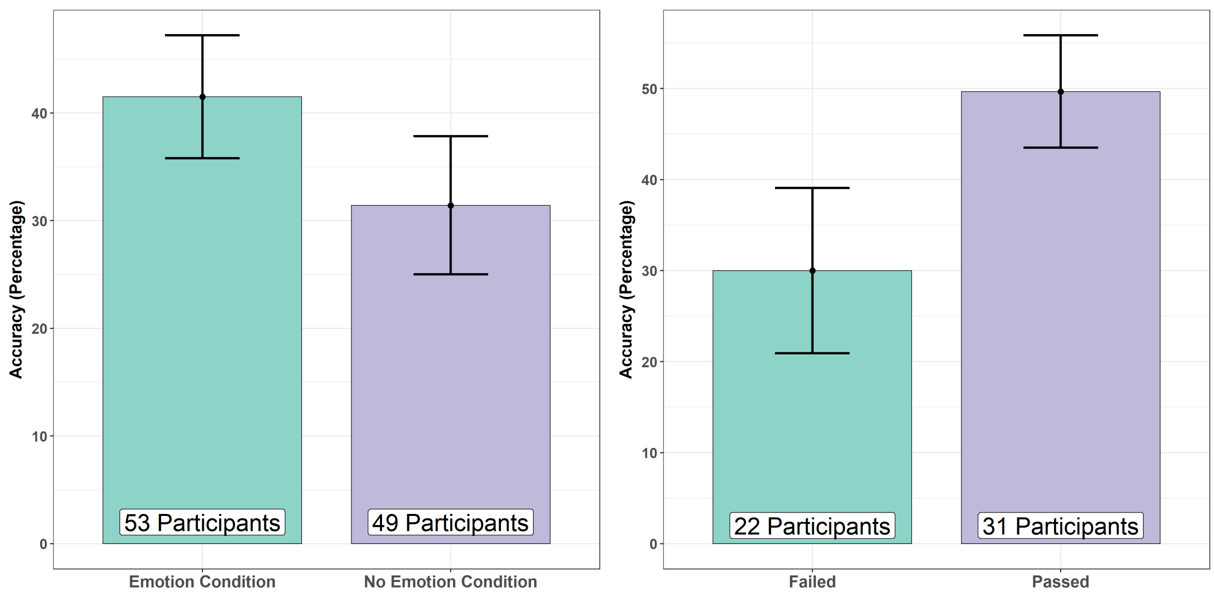


Figure 5.9: Emotions that participants who failed in the training could not learn versus the average number of attempts each participant had to guess particular emotion.

Perception Accuracy

Participants in emotion condition were significantly better than participants in no emotion condition to predict scenarios shown during the main task ($se = 0.04, t = 2.287, p = .024$). Results of LMM analysis for the model with minimum AIC can be seen in Table 5.5.

The average perception accuracy of participants in both conditions was compared in Figure 5.10. In addition to LMM analysis, statistical differences between two groups were calculated using Welch's two sample t-test [48]. Participants in the no emotion condition had an average perception accuracy of around 30%, whereas participants in the emotion condition had significantly higher average perception accuracy of around 40% (t=2.3573, df = 97.623, p=0.0204). Participants in the emotion condition were also divided into two depending on their success during the training. The ones who failed in the training step had a mean accuracy of $\approx$ 30% while those who succeeded in the training step had significantly higher mean accuracy of $\approx$ 50% (t=-3.7065, df=39.6, p=0.00064).



(a) Comparing two different conditions

(b) Emotion condition based on training success

Figure 5.10: Average perception accuracy of participants compared for each condition given on the left (significantly different than each other t=2.3573, df = 97.623, p=0.0204) and average perception accuracy of participants in emotion condition compared based on their success during the emotion training step (significantly different than each other t=-3.7065, df=39.6, p=0.00064).

Average perception accuracy of participants in no emotion condition based on whether they failed during the training step was compared in Figure 5.11, and it was discovered that they had a similar mean accuracy around 30% to guess given SAR scenarios. Accuracy of these two subgroups in no emotion condition was not used individually in the previous

Figure 5.10 for comparison with the ones in emotion condition since participants in no emotion condition did not see emotional lights during the main task.

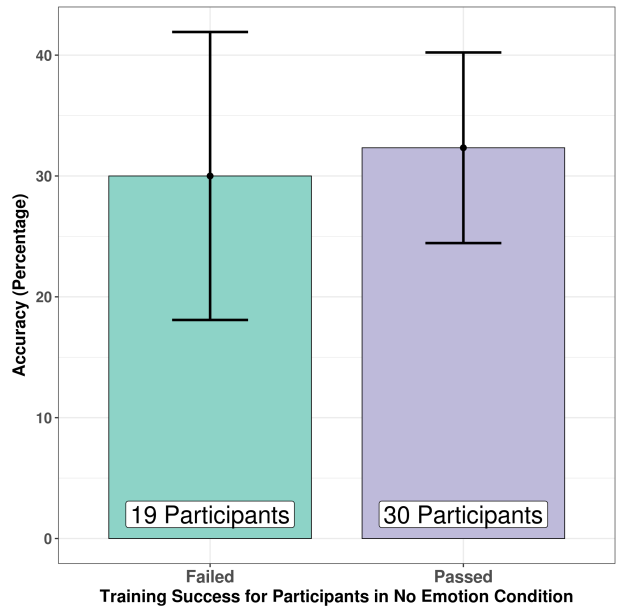
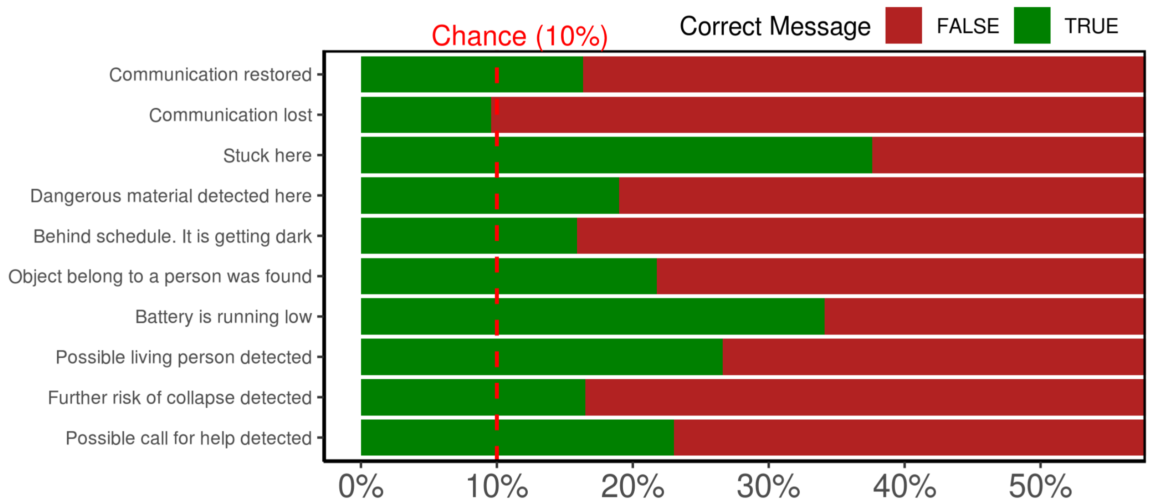


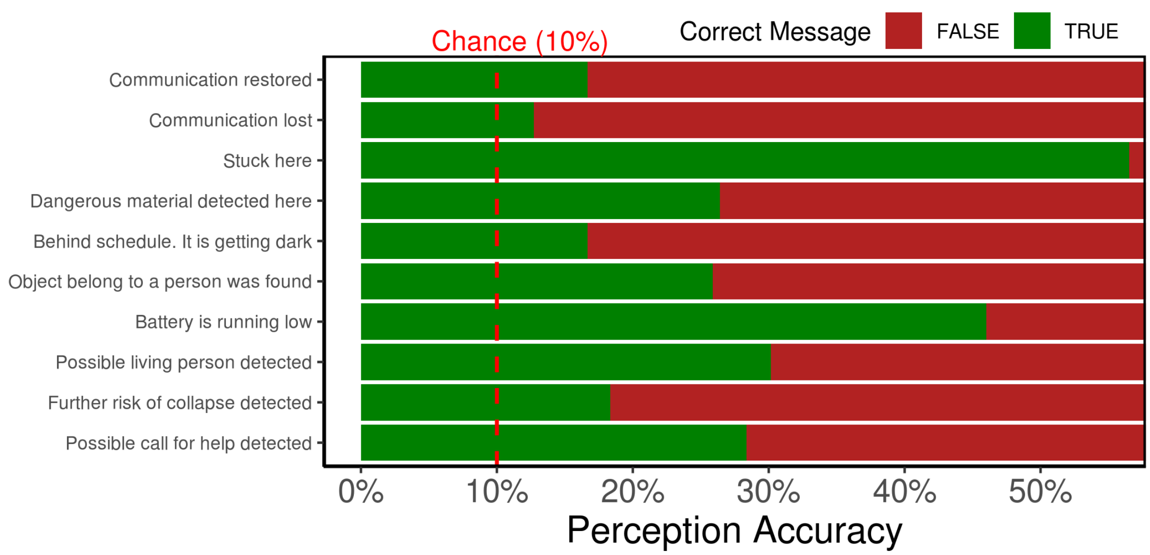
Figure 5.11: Average perception accuracy of participants for only no emotion condition. Participants divided into two based on their success during the emotion training step: 30 participants passed the training (i.e., learned all emotions successfully) and 19 participants failed. There is no significant difference between these two groups.

The success in recognition of individual SAR scenarios was also analyzed for three groups: (a) participants in emotion condition including both the ones who failed during the training and who passed the training step, (b) participants in emotion condition who passed the training and (c) participants in no emotion condition including both the ones failed and passed the training step (see Figure 5.12). Perception accuracy in the group (b) is the largest, whereas it is the lowest in group (c). None of the scenarios was recognized with more than 60% accuracy. In all cases, scenarios “Stuck here” and “Battery is running low” were identified with the highest accuracy than others.

The success in recognition of individual SAR scenarios regarding their type (positivity vs. negativity) was also examined. This analysis was done for three participant groups: (a) participants in emotion condition including both the ones who failed during the training and who passed the training step, (b) participants in emotion condition who passed the training, and (c) participants in no emotion condition including both the ones failed and

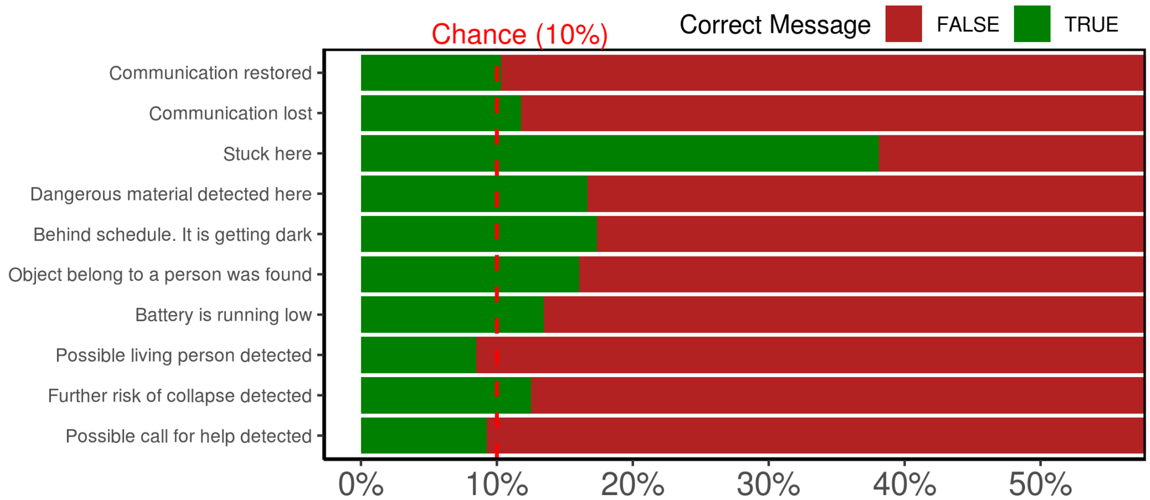


(a) Emotion Condition (53 Participants, including participants who failed in the training step)



(b) Emotion Condition (31 Participants, only including those who passed the training step)

Figure 5.12: Perception accuracy of SAR-related messages shown the participants during the experiment 3. Messages shown to participants are not the ones in these plots. The shortened version of the original messages was used here for visualization purposes. The horizontal axis is also limited to 50% for the same reason. The figure continues on the next page...



(c) No emotion condition

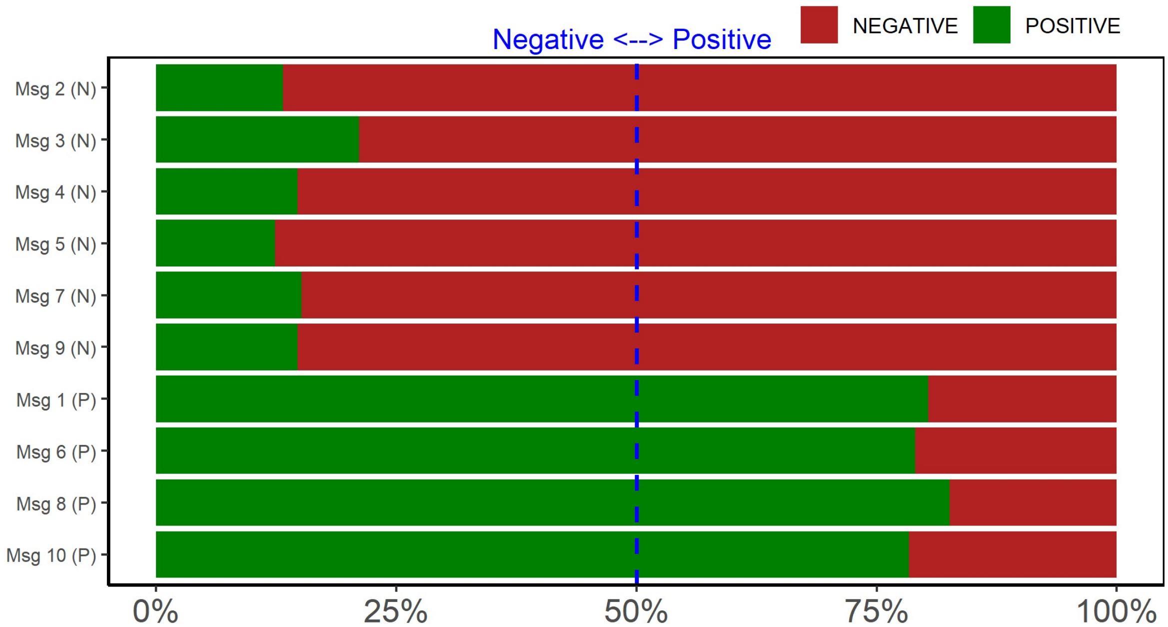
Figure 5.12: Perception accuracy of SAR-related messages shown the participants during the experiment 3. Messages shown to participants are not the ones in these plots. The shortened version of the original messages was used here for visualization purposes. The horizontal axis is also limited to 50% for the same reason.

passed the training step (see Figure 5.13). Participants in group (b) were the best to understand whether the scenario was positive or negative (over 90% accuracy), whereas participants in group (c) were the worst (for some scenarios, their accuracy is less than the chance, i.e., 50%).

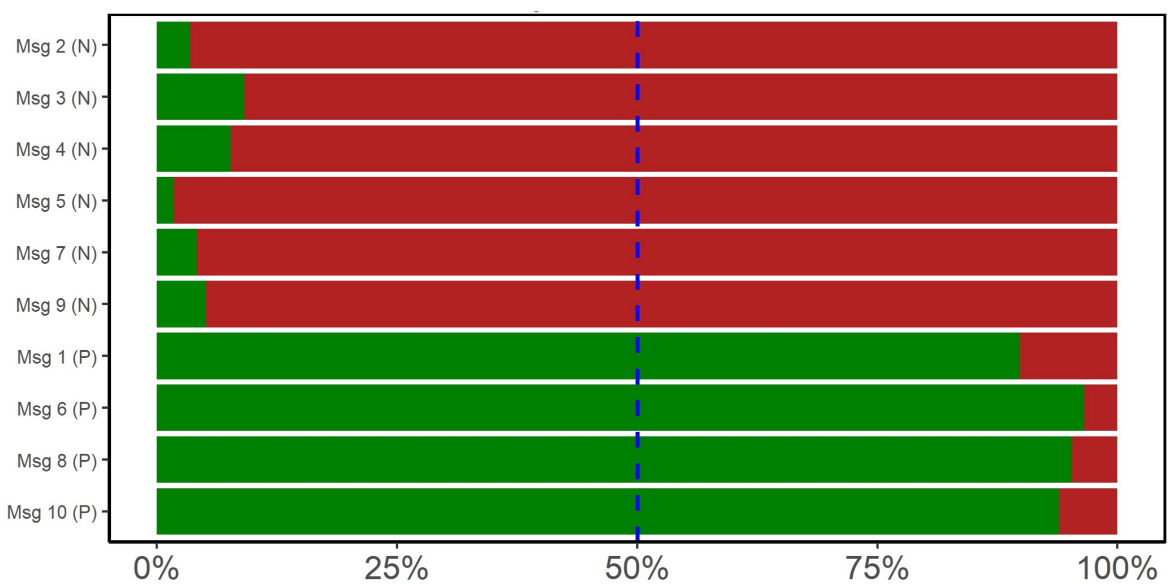
Participants' responses to shown scenarios were analyzed to understand which scenarios were confused with each other. Analysis was again done for three groups: (a) participants in emotion condition, including both the ones who failed during the training and who passed the training step, (b) participants in emotion condition who passed the training, and (c) participants in no emotion condition including both the ones failed and passed the training step (see Table 5.8).

Response Time

Despite the significant difference in perception accuracy of participants depending on the condition they were assigned to, there was no significant difference regarding their response time, i.e., the time it took participants to guess the shown scenarios ($se = 6.06, t = -0.05, p = .96$). Statistical findings of LMM analysis can be seen in Table 5.6.

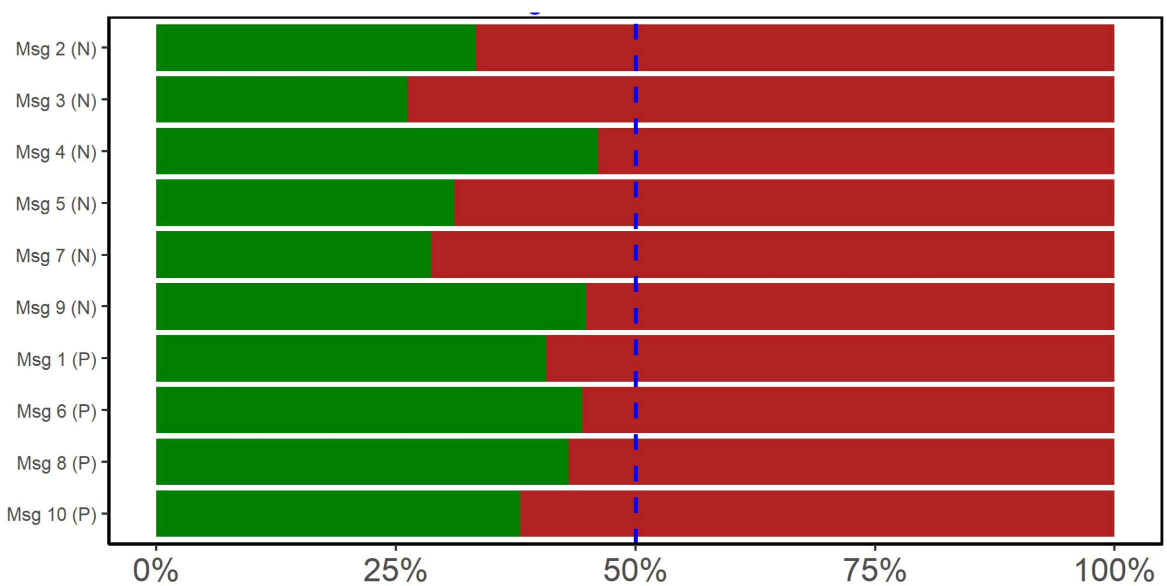


(a) Emotion Condition (53 Participants, including participants who failed in the training step)



(b) Emotion Condition (31 Participants, only including those who passed the training step)

Figure 5.13: Perception accuracy of SAR-related messages shown the participants during experiment 3 regarding their type (i.e., positivity/negativity). See Table 5.1 for the scenario names.



(c) No emotion condition

Figure 5.13: Perception accuracy of SAR-related messages shown the participants during experiment 3 regarding their type (i.e., positivity/negativity). See Table 5.1 for the scenario names.

Table 5.8: Participants' responses to shown scenarios during the main task. Rows show correct scenario numbers while columns show participants' selections. Scenarios selected significantly more than others were calculated via binomial tests and shown in pink while scenarios selected more than a chance shown in bold (***: $p < .001$, **: $p < .01$, and *: $p < .05$). Positive scenarios are shown with green while negative ones are shown with red.

(a) Participants in emotion condition including the ones failed in the training step (53 in total)

	Responses	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Not Sure
Correct Scenarios	One	**17***	2	4	4	4	**26*****	3	**24*****	3	15	2
	Two	1	9	14	**16***	10	5	12	3	**18****	3	3
	Three	6	11	**35*****	10	2	4	8	5	5	4	3
	Four	6	9	9	**19****	11	4	**17***	2	**16***	2	5
	Five	5	16	13	**19****	**17***	5	12	3	15	0	2
	Six	8	3	5	6	3	**22*****	1	**30*****	3	**19****	1
	Seven	4	5	13	8	9	3	**30*****	1	6	5	4
	Eight	13	3	5	5	2	**26*****	3	**29*****	1	**22****	0
	Nine	4	14	10	**16***	9	3	**16***	2	**16***	5	2
	Ten	12	2	5	4	7	**20****	2	**29*****	4	**26*****	2

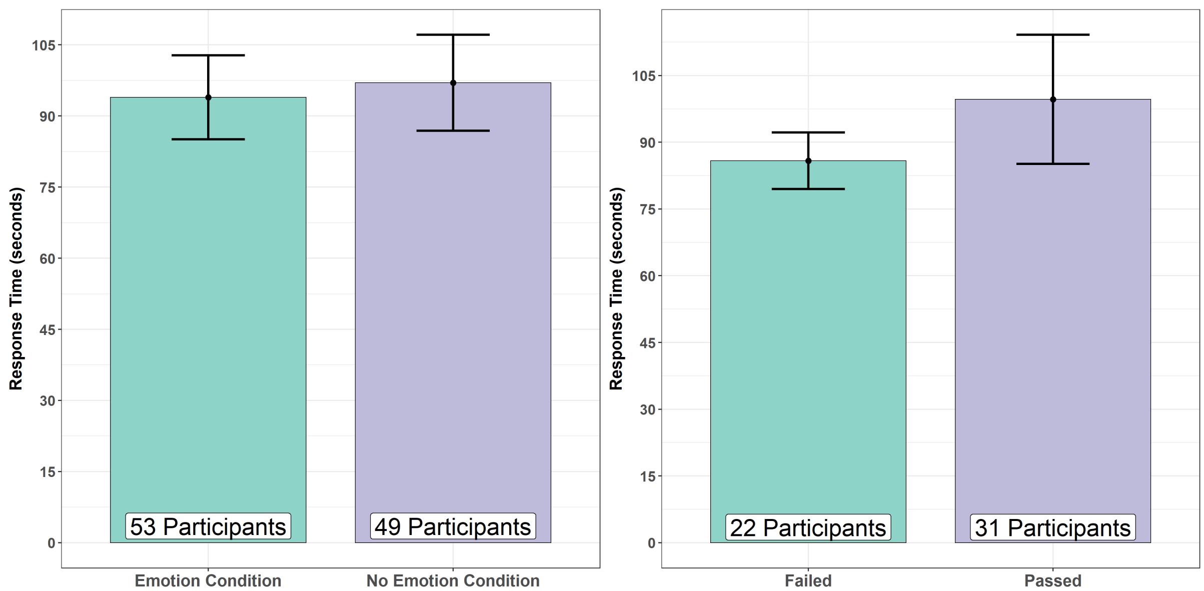
(b) Participants in Emotion Condition who passed the training step (31 Participants in Total)

	Responses	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Not Sure
Correct Scenarios	One	10	0	1	1	1	**17*****	1	**16*****	2	10	1
	Two	0	7	9	9	5	0	6	0	**15*****	2	2
	Three	1	6	**26*****	3	2	1	2	1	1	1	2
	Four	1	4	3	**14*****	6	0	6	0	**15*****	0	4
	Five	2	**12***	8	10	10	0	6	0	10	0	2
	Six	5	0	0	1	0	**15*****	0	**22*****	2	**13****	0
	Seven	1	4	**9***	4	4	0	**23*****	0	2	1	2
	Eight	8	0	0	1	0	**17*****	2	**19*****	0	**16*****	0
	Nine	1	8	6	**14****	5	1	**12***	0	**11***	1	1
	Ten	8	1	2	0	1	**13***	0	**21*****	0	**19*****	2

(c) Participants in No Emotion Condition (49 Participants in Total)

	Responses	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Not Sure
Correct Scenarios	One	10	4	5	11	13	10	8	12	12	5	7
	Two	7	12	13	12	12	7	6	7	9	11	6
	Three	7	4	**32*****	6	6	7	9	5	2	2	4
	Four	12	6	7	**18***	6	12	5	12	13	11	6
	Five	9	8	5	8	**17****	8	**15***	6	9	5	8
	Six	6	7	8	15	11	**18***	8	13	11	11	4
	Seven	9	12	12	12	13	10	14	5	8	5	4
	Eight	10	6	11	14	7	14	8	9	11	10	6
	Nine	7	10	8	12	9	13	6	**17***	14	11	5
	Ten	10	8	9	14	10	7	9	11	12	10	8

Figure 5.14 compares mean response time of participants for each condition. Participants in the no emotion condition had a slightly more response time than participants in the emotion condition, but as explained before, this difference was not significant. Participants in the emotion condition were divided depending on their training success. It took less time to respond for participants who failed during the training step, but no significant difference was detected. Overall, the average response time is around 90 seconds, including the videos' length (the first video is 36 seconds on average and the second video is 18 seconds on average) and excluding the idle time that participants did not spend on the main task.



(a) Comparing two different conditions

(b) Emotion condition based on training success

Figure 5.14: Average response time of participants compared for each condition given on the left (no significant difference) and average response time of participants in emotion compared based on their success during the emotion training step (no significant difference).

Participants in the no emotion condition were also analyzed regarding their average response time. Participants who failed to learn all the emotions during the training step took a slightly longer time to guess the SAR scenarios, but this difference is not significant as it can be seen in Figure 5.15.

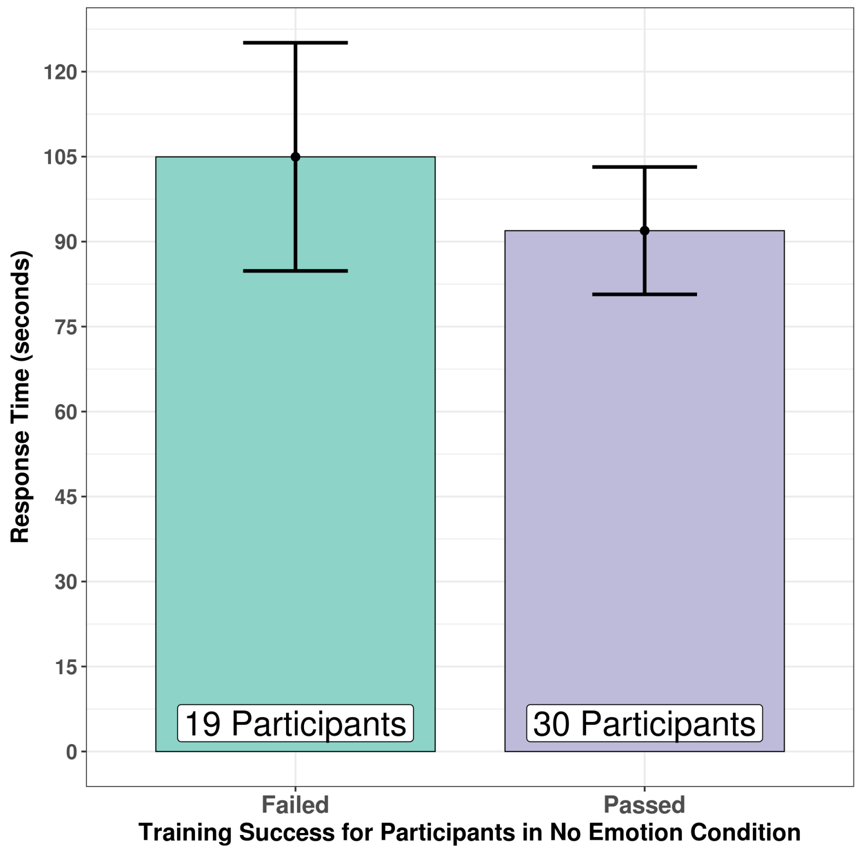


Figure 5.15: Average response time of participants for only no emotion condition. Participants divided into two based on their success during the emotion training step: 30 participants passed the training, (i.e., learned all emotions successfully), and 19 participants failed

Questionnaire Results

Participants in both conditions were asked to report how hard it was for them to read the text messages on a continuous scale where 0 corresponds to not hard at all and 1000 corresponds to very hard. They were also given an option to state that they could not read the text at all. As it can be seen in Figure 5.16, the majority of the participants found the shown text messages very difficult to read, with 45 of them stating that they could not read them. To test the relation between their response to this question and their performance during the main task, reported noise levels were factorized into five, with noise level 1 represents their responses between 0-250 and level 5 represents they could not read the text. LMM analysis did not result in any significance relation between reported noise level and performance metrics perception accuracy ($se = 1.868e - 2, t = -0.52, p = 0.60435$) & response time ($se = 2.81808, t = -1.204, p = 0.23127$).

Participants in emotion condition were asked to report how much they agree with the statement "I think the robot's use of lights in order to convey emotions was helpful to guess messages sent by the robot", and participants in no emotion condition were asked to report how much they agree with the statement "I would prefer the robot to use lights in order to convey emotions which might be helpful to guess messages sent by the robot". Their responses to both statements were shown in Figure 5.17. For both conditions, the mean value is around 750 (1000 corresponds, I totally agree), stating that most of the participants either prefer the robot to use lights or found the robot's use of lights helpful depending on the condition.

A linear mixed-effect model was fit for participants' data in the emotion condition to predict response time and perception accuracy based on their ratings of how helpful they found the lights. While there was no significant effect of ratings on perception accuracy (se = 1.514e-04, t = 0.819, p = 0.4131), there was a significant negative effect of how helpful they found the lights on their response time (se = 2.313e-02, t = -2.479, p = 0.01639) indicating that the more useful the participants rated the emotions, the less time they spent on guessing the messages (see Figure 5.18).

5.3.6 Discussion

The motivation of this experiment was to investigate whether the usage of emotions can complement the communication in robot-assisted SAR teams to improve situational awareness of rescue workers. It was discovered that a rescue robot that uses emotions to complement the conveyed message has the potential to increase the accuracy of understanding

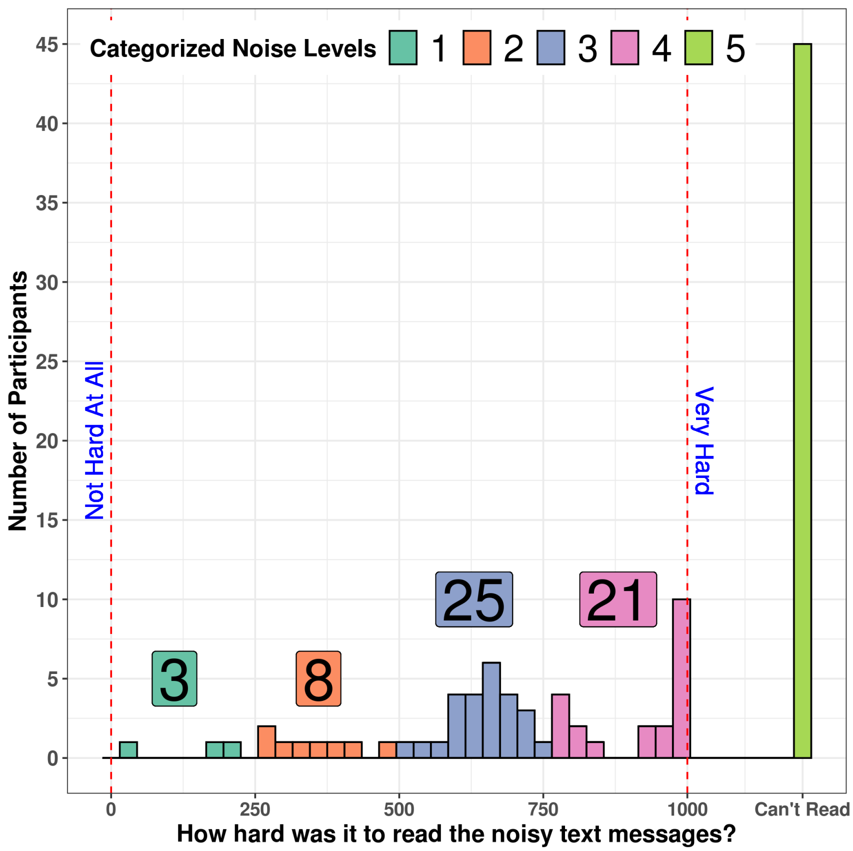
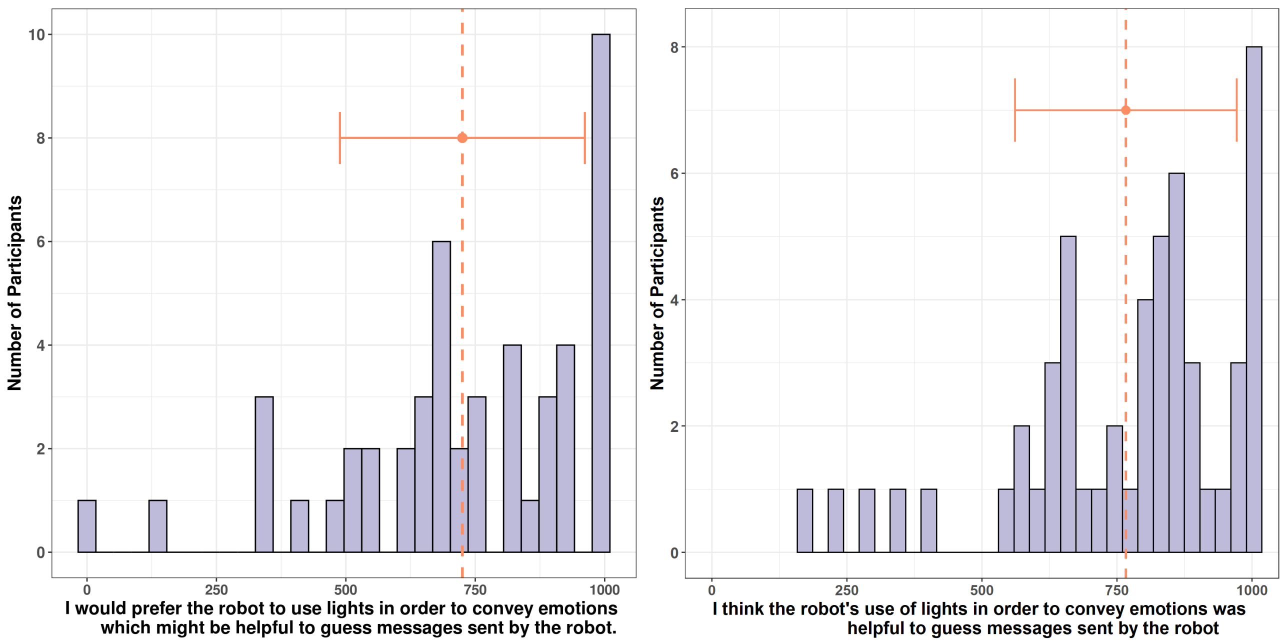


Figure 5.16: Participants' response to the survey question "How hard was it to read the noisy text messages?".



(a) No Emotion Condition (b) Emotion Condition

Figure 5.17: Response of participants in both conditions. Mean and standard deviation were shown with orange dashed and solid lines respectively.

the messages, while it did not have any significant effect on the time it took participants to guess the scenario, i.e., response time. It is important to highlight that this experiment was conducted with the participation of people through crowd-sourcing, so the findings of this experiment need to be further tested with actual rescue workers. In the following subsections, the findings of the experiment are discussed comprehensively.

Design of Affective Lights

All implemented emotions (tired, happy, fear, excited, and annoyed) were perceived successfully with the accuracy of more than 50% (chance is 20%) after training (see Figure 5.8). Among these emotions, the positive emotions happy and excited were recognized more accurately than the negative emotions tired, fear, and annoyed. The majority of participants who failed in the training step had difficulty distinguishing the negative emotions accurately (see Table 5.7). Fear was the most commonly confused emotion among the negative emotions, which was confused with either annoyed or tired. Furthermore, fear had the

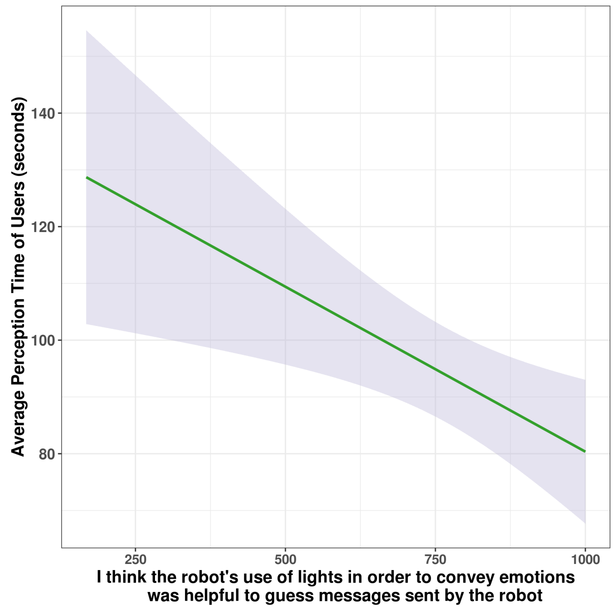


Figure 5.18: Average response time of each user vs how useful they found lights to guess messages sent by the robot

largest average number of attempts per participant shown in Figure 5.9. The reason behind this misrecognition may be that EPA values for negative emotions are close to each other. The difference in the Potency (P) dimension may not be recognized well solely based on changes in light intensity since the experiment was not conducted in person. Regardless, the high rate of failure in training due to the design of affective expressions poses the question of their intuitiveness. Employing a different way of matching lights parameters with EPA dimensions and/or having additional parameters (like using different patterns for each emotion) might improve the perception accuracy of emotions. However, it is not possible to draw further conclusions with the data obtained.

Perception Accuracy

Findings of this experiment showed that participants in the emotion condition had a better perception accuracy in guessing shown SAR scenarios. Their perception accuracy increased with their training success. Moreover, it was also detected that affective lights are a good way to inform participants about the positivity/negativity of the scenario (see Figure 5.13). All these findings support our hypothesis **H1** where we suggested that participants in the emotion condition will have better situational awareness, i.e., their understanding of what is happening in the disaster area.

Response Time

Although the findings of this experiment supported the hypothesis **H1**, they did not provide enough information to either confirm or reject hypothesis **H2**, where we suggested that a rescue robot with an ability to communicate through affective expressions will decrease participants' response time by complementing other communication modalities. We could not find any significant difference in response time between participants in the emotion condition and the no emotion condition.

The majority of participants could not read the noisy text messages shown during the main task, and they found the noise level quite high to read, as shown in Figure 5.16. This is the reason that the insignificance in response time between participants in emotion and no emotion condition can neither confirm nor reject hypothesis **H2**. In order to evaluate **H2**, affective expressions should complement other modalities (text), but they were simulated as failed in this experiment since the majority of the participants could not read the messages.

5.4 Limitations and Future Work

The first limitation of this study is caused by the online nature of its design. Participants were shown either simulated or real videos of the Husky robot, and they were asked to guess what is happening in the disaster area, considering the movements of the robot, affective lights (in emotion condition), and the noisy text message received. Nonetheless, they did not experience in-person interaction with the robot or the real atmosphere of the disaster area, which might change their situation and findings of this experiment.

The second limitation of this study is the design of affective lights. Participants could not distinguish negative emotions as well as positive emotions. Although they differ in real life, this difference is not that clear in the recorded videos due to the technical difficulties of recording high-speed, low/high brightness of LED lights. An additional future study with real interaction can improve the accuracy of emotion perception. It may also improve the findings of this study since participants' success in perceiving SAR scenarios increases as their training success increases (see Figure 5.10).

Results of this experiment did not provide enough information to either confirm or reject hypothesis **H2** since the majority of the participants found the noisy text messages difficult to read. In the future, it would be a good idea to conduct another experiment with a similar setup but with a lower noise level in the text messages. Such an experiment might help us to evaluate **H2** better.

Design of SAR scenarios might affect the findings of this experiment. Despite the fact that all SAR scenarios were perceived with accuracy more than the chance, there is still room to design better scenarios to strengthen simulated scenarios. Another experiment with the same setup but different simulated scenario patterns can be conducted as future work to check whether similar results regarding affective lights can be obtained.

In the future, other experiments with different rescue robots should be conducted to study the effect of robot priming. The ability to express emotions can be implemented on other robots easily in a similar manner. Such an approach might be needed since, despite rescue robots being less anthropomorphic, their appearance can still affect participants' success in understanding the disaster scene. Furthermore, other communication modalities such as voice can also be integrated into the current communication system to study the effect of multi-modal communication and affective expressions.

Lastly, the majority of the participants were not familiar with SAR. As a future experiment, the findings of this study might be compared with a similar study conducted with the participation of real SAR workers. Their experience with SAR missions might differ from the findings of this study.

5.5 Conclusion

- For this experiment, affective expressions were designed using lights based on EPA dimensions suggested by ACT and implemented on the appearance-constrained SAR robot Husky.
- Common situations occurring during SAR missions (the ones from the first experiment) were simulated in Gazebo, and emotions corresponding to these situations were used during the study.
- Text messages were chosen as an additional communication channel, and its failure was simulated via noise that made messages very difficult to read.

- Considering noisy text message, movements of Husky in the simulation (and affective expressions in the emotion condition), participants were asked to guess what message Husky wanted to convey.
- Findings suggested that participants in the emotion condition had significantly higher perception accuracy in understanding the messages sent by Husky. At the same time, there was no significant difference in their response time to guess the messages.

Chapter 6

Conclusion

6.1 Summary of Findings

This thesis proposes bringing affective expressions into robot-assisted SAR to increase the efficiency of human-robot communication. In this regard, the following research questions were answered:

RQ1 Is there a consensus on what emotions to be used by USAR robots when they try to convey information about the situations commonly occurring during USAR operations?

- Based on the literature review, common situations happening during USAR missions were found (see Table 3.1). We conducted an online study and asked participants to match these scenarios with emotions (Chapter 3). There was consensus in participants' responses. This resulted in a mapping between the scenarios and emotions as presented in Table 3.3.

RQ2 Is the mapping between emotions and USAR situations robust and not dependent on the wording of the sentences?

- In the first experiment presented in Chapter 3, participants were divided into two conditions based on the wording of USAR related scenarios. The resulted mappings between the scenarios and emotions were similar between two conditions suggesting that the mapping was robust (see Table 3.3).

RQ3 How can a mapping between SAR related sentences and emotions be obtained, and is there a way to generalize such mapping without limiting it to a specific set of emotions?

- This research question resulted from the methodology of the first experiment, where participants were provided with a list of emotions to select (Chapter 3). In the second experiment (Chapter 4), they were asked to rate EPA dimensions in ACT. Afterward, their ratings were processed to find the associated emotions. When the same emotion set as in the first experiment was used during the post-processing, the same mapping was obtained between common scenarios and emotions as given in Table 4.4. This finding answered this research question indicating that it is possible to obtain the mapping using EPA dimensions without a need to limit the emotion set. Therefore, a robot with the capability to show a different emotion set (due to its embodiment) will be able to use the same mapping.

RQ4 How can affective expressions be designed and implemented on appearance constrained SAR robots using lights?

- We designed affective expressions using lights based on ACT. EPA dimensions were transformed into light parameters as in Table 5.2. Five affective expressions that were found in the mapping obtained during the first experiment (Chapter 3) were implemented on appearance constrained Clearpath's Husky robot. Participants had high accuracies in recognizing these emotions after training (see Figure 5.8).

RQ5 Can affective expressions complement and improve multi-modal communication in human-robot SAR teams?

- Participants in the third experiment (Chapter 5) were divided into two conditions emotion versus no emotion. They both tried to guess what is happening in the disaster area in SAR situation where they can not read the text message sent by the robot. Although participants in both conditions had similar response time to understand given scenarios (see Figure 5.14), participants in emotion condition had significantly higher perception accuracy (see Figure 5.10). These findings suggest that bringing affective expressions into SAR context has the potential to improve communication in robot-assisted SAR teams. It is important to highlight that participants of this study were not quite familiar with SAR. They were recruited via crowd-sourcing due to the COVID-19 pandemic.

6.2 Contribution to Knowledge

Research presented in this thesis had the following contributions to robot-assisted SAR and HRI:

- We identified the most common situations happening during USAR tasks and introduced the idea to use affective expressions to complement multi-modal communication in robot-assisted SAR teams. A similar approach can be applied to improve HRI in other contexts by obtaining a mapping between specific context-related scenarios and emotions like the one in Table 3.3. In addition, we took advantage of ACT and EPA dimensions to automate the process of obtaining a mapping and allowing post-processing of data. Usage of EPA ratings in such a way may also contribute to the generalizability of the idea of using affective expressions to improve human-robot communication.

- We designed affective expressions based on EPA dimensions and implemented them on an appearance constrained robot Husky using LED lights. This method produces a nice way to implement emotions quantitatively thanks to ACT. The usage of simple LED strips to achieve this offers an affordable way to create affective appearance constrained robots for HRI researchers.

- We showed the increase in participants' situational awareness when interacting with an affective rescue robot capable of expressing emotions. This finding may contribute to building more social rescue robots in the future.

6.3 Limitations and Future Work

The major limitation of this thesis is about conducting online studies. Due to the online nature of these studies, participants did not have a chance to interact with real SAR robots. They also did not experience a real-life SAR scenario, which could affect the obtained results. Despite these limitations, such an online approach was followed as the first step for this direction of research to not bias participants with the appearance of a particular robot, and it also helped with reducing the experimenter bias [70]. Furthermore, this approach has been shown to be effective in many HCI and HRI studies and has gained more attention since COVID-19 has affected the feasibility of conducting in-person HRI studies. Nonetheless, future work is needed to investigate if/how the findings of this thesis would translate to real-life situations.

When affective rescue robots are used in real SAR missions, there might be other challenges regarding the usage of affective expressions in the disaster area. For example, perception of affective expressions might differ in environments with varying visibility conditions such as smoke, rain, or dark as we investigated in [64]. While we showed that recognition of affective expressions conveyed through a robot's body and head gestures could be robust, to a reasonable extent, under different visibility constraints [64], future work is needed to examine the effects of visibility conditions on the accuracy of recognition of affective expressions. How perception accuracy of affective expressions would be affected by other factors such as culture, gender, and selected modality of expressing emotions needs to be carefully examined in future as well.

Participants recruited for all the experiments presented in this thesis were not familiar with SAR. It is unclear what would happen to the findings of this thesis if participants had an experience with SAR situations. In the future, further studies with people who are more familiar with SAR and rescue workers themselves need to be recruited to confirm the findings of this thesis.

References

[1] Annette L Adams, Terri A Schmidt, Craig D Newgard, Carol S Federiuk, Michael Christie, Sean Scorvo, and Melissa DeFreest. Search is a time-critical event: when search and rescue missions may become futile. *Wilderness & Environmental Medicine*, 18(2):95–101, 2007.

[2] Federal Emergency Management Agency and U.S.F. Administration. *Technical Rescue Program Development Manual*. CreateSpace Independent Publishing Platform, 2013.

[3] Sami Alperen Akgun, Moojan Ghafurian, Mark Crowley, and Kerstin Dautenhahn. Using emotions to complement multi-modal human-robot interaction in urban search and rescue scenarios. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, ICMI '20, page 575–584, New York, NY, USA, 2020. Association for Computing Machinery.

[4] Sami Alperen Akgun, Moojan Ghafurian, Mark Crowley, and Kerstin Dautenhahn. Integrating affective expressions into the search and rescue context in order to improve non-verbal human-robot interaction. In *Exploring Applications for Autonomous Non-Verbal Human-Robot Interactions Workshop of ACM/IEEE International Conference on Human-Robot Interaction*, 2021.

[5] Mohamad K Allouche and Abdeslem Boukhtouta. Multi-agent coordination by temporal plan fusion: Application to combat search and rescue. *Information Fusion*, 11(3):220–232, 2010.

[6] Rebecca Andreasson, Beatrice Alenljung, Erik Billing, and Robert Lowe. Affective touch in human–robot interaction: conveying emotion to the nao robot. *International Journal of Social Robotics*, 10(4):473–491, 2018.

[7] Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Kary Främling. Explainable agents and robots: Results from a systematic literature review. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19, page 1078–1088, Richland, SC, 2019. International Foundation for Autonomous Agents and Multiagent Systems.

[8] Theo Araujo. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85:183 – 189, 2018.

[9] Todd W Arnold. Uninformative parameters and model selection using akaike's information criterion. *The Journal of Wildlife Management*, 74(6):1175–1178, 2010.

[10] Michael Baker, Robert Casey, Brenden Keyes, and Holly A Yanco. Improved interfaces for human-robot interaction in urban search and rescue. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, volume 3, pages 2960–2965. IEEE, 2004.

[11] Kim Baraka, Ana Paiva, and Manuela Veloso. Expressive lights for revealing mobile service robot state. In *Robot 2015: Second Iberian Robotics Conference*, pages 107–119. Springer, 2016.

[12] Christoph Bartneck, Andreas Duenser, Elena Moltchanova, and Karolina Zawieska. Comparing the similarity of responses received from studies in amazon mechanical turk to studies conducted online and with direct recruitment. *PLOS ONE*, 10(4):1–23, 04 2015.

[13] Douglas Bates. Fitting linear mixed models in r. *R news*, 5(1):27–30, 2005.

[14] Aryel Beck, Lola Cañamero, Antoine Hiolle, Luisa Damiano, Piero Cosi, Fabio Tesser, and Giacomo Sommavilla. Interpretation of emotional body language displayed by a humanoid robot: A case study with children. *International Journal of Social Robotics*, 5(3):325–334, 2013.

[15] Jenay M Beer, Akanksha Prakash, Tracy L Mitzner, and Wendy A Rogers. Understanding robot acceptance. Technical report, Georgia Institute of Technology, 2011.

[16] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[17] Mel V Bennett and Iain Matthews. Life-saving uncooled ir camera for use in fire-fighting applications. In *Infrared Technology and Applications XXII*, volume 2744, pages 549–554. International Society for Optics and Photonics, 1996.

[18] Markus Bernard, Konstantin Kondak, Ivan Maza, and Anibal Ollero. Autonomous transportation and deployment with aerial robots for search and rescue missions. *Journal of Field Robotics*, 28(6):914–931, 2011.

[19] Cindy L Bethel and Robin R Murphy. Affective expression in appearance constrained robots. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 327–328, 2006.

[20] Cindy L. Bethel and Robin R. Murphy. Non-facial and non-verbal affective expression for appearance-constrained robots used in victim management*. *Paladyn, Journal of Behavioral Robotics*, 1(4):219–230, 2010.

[21] Sandra Leanne Bosacki and Chris Moore. Preschoolers' understanding of simple and complex emotions: Links with gender and language. *Sex roles*, 50(9-10):659–675, 2004.

[22] Cynthia Breazeal. Emotion and sociable humanoid robots. *International journal of human-computer studies*, 59(1-2):119–155, 2003.

[23] EL Buckland, HA Volk, CC Burn, and SM Abeyesinghe. Owner perceptions of companion dog expressions of positive emotional states and the contexts in which they occur. *Animal Welfare*, 23(3):287–296, 2014.

[24] John G Buckley, Karen J Heasley, Pete Twigg, and David B Elliott. The effects of blurred vision on the mechanics of landing during stepping down by the elderly. *Gait & posture*, 21(1):65–71, 2005.

[25] Charlotte C Burn. Bestial boredom: A biological perspective on animal boredom and suggestions for its scientific investigation. *Animal Behaviour*, 130:141–151, 2017.

[26] Lola Canamero. Building emotional artifacts in social worlds: Challenges and perspectives. *Emotional and intelligent II: the tangled knot of social cognition*, 2001.

[27] Walter B Cannon. The james-lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4):106–124, 1927.

[28] Jennifer Casper and Robin R. Murphy. Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 33(3):367–385, 2003.

[29] Jennifer L Casper, Mark Micire, and Robin R Murphy. Issues in intelligent robots for search and rescue. In *Unmanned ground vehicle technology II*, volume 4024, pages 292–302. International Society for Optics and Photonics, 2000.

[30] Justine Cassell and Timothy Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modelling and User-Adapted Interaction*, 13, 04 2001.

[31] Carlos Mauricio Castaño Díaz. Defining and characterizing the concept of internet meme. *Ces Psicología*, 6(2):82–104, 2013.

[32] Patricia Chaffey, Ron Artstein, Kallirroi Georgila, Kimberly A Pollard, and Setareh Nasihati Gilani. Developing a virtual reality wildfire simulation to analyze human communication and interaction with a robotic swarm during emergencies. In *Workshop on Human Language Technologies in Crisis and Emergency Management*, 2019.

[33] Tathagata Chakraborti, Subbarao Kambhampati, Matthias Scheutz, and Yu Zhang. AI challenges in human-robot cognitive teaming. *CoRR*, abs/1707.04775, 2017.

[34] Wang Chang, Wu Lizhen, Yan Chao, Wang Zhichao, Long Han, and Yu Chao. Coactive design of explainable agent-based task planning and deep reinforcement learning for human-uavs teamwork. *Chinese Journal of Aeronautics*, 2020.

[35] Albert Y Chen, Feniosky Peña-Mora, Albert P Plans, Saumil J Mehta, and Zeeshan Aziz. Supporting urban search and rescue with digital assessments of structures and requests of response resources. *Advanced Engineering Informatics*, 26(4):833–845, 2012.

[36] Andry Chowanda, Martin Flintham, Peter Blanchfield, and Michel Valstar. Playing with social and emotional game companions. In *International Conference on Intelligent Virtual Agents*, pages 85–95. Springer, 2016.

[37] Nikhil Churamani, Francisco Cruz, Sascha Griffiths, and Pablo Barros. icub: learning emotion expressions using human reward. *arXiv preprint arXiv:2003.13483*, 2020.

[38] Emily C Collins, Tony J Prescott, and Ben Mitchinson. Saying it with light: A pilot study of affective communication using the miro robot. In *Conference on Biomimetic and Biohybrid Systems*, pages 243–255. Springer, 2015.

[39] Emily C Collins, Tony J Prescott, Ben Mitchinson, and Sebastian Conran. Miro: a versatile biomimetic edutainment robot. In *Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology*, pages 1–4, 2015.

[40] Lukáš Danev, Marten Hamann, Nicolas Fricke, Tobias Hollarek, and Dennys Paillacho. Development of animated facial expressions to express emotions in a robot: Roboticon. In *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, pages 1–6. IEEE, 2017.

[41] Quoc Khanh Dang and Young Soo Suh. Human-following robot using infrared camera. In *2011 11th International Conference on Control, Automation and Systems*, pages 1054–1058. IEEE, 2011.

[42] Per-Erik Danielsson. Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248, 1980.

[43] C Darwin. The expression of emotion in man and animals. london, england: Murray, 1872.

[44] Kerstin Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical transactions of the royal society B: Biological sciences*, 362(1480):679–704, 2007.

[45] Kerstin Dautenhahn, Alan H Bond, Lola Canamero, and Bruce Edmonds. *Playing the emotion game with Feelix: What can a LEGO robot tell us about emotion?*, volume 3. Springer Science & Business Media, 2002.

[46] Kerstin Dautenhahn, Chrystopher L Nehaniv, Michael L Walters, Ben Robins, Hatice Kose-Bagci, N Assif Mirza, and Mike Blow. Kaspar–a minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics*, 6(3-4):369–397, 2009.

[47] Geert De Cubber, Haris Balta, and Claude Lietart. Teodor: A semi-autonomous search and rescue and demining robot. In *Applied Mechanics and Materials*, volume 658, pages 599–605. Trans Tech Publ, 2014.

[48] Marie Delacre, Daniël Lakens, and Christophe Leys. Why psychologists should by default use welch's t-test instead of student's t-test. *International Review of Social Psychology*, 30(1), 2017.

[49] Jeffrey Delmerico, Stefano Mintchev, Alessandro Giusti, Boris Gromov, Kamilo Melo, Tomislav Horvat, Cesar Cadena, Marco Hutter, Auke Ijspeert, Dario Floreano, et al. The current state and future outlook of rescue robotics. *Journal of Field Robotics*, 36(7):1171–1191, 2019.

[50] Daantje Derks, Arjan ER Bos, and Jasper Von Grumbkow. Emoticons and online message interpretation. *Social Science Computer Review*, 26(3):379–388, 2008.

[51] Sidney D'Mello and Rafael A. Calvo. Beyond the basic emotions: What should affective computing compute? In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, page 2287–2294, New York, NY, USA, 2013. Association for Computing Machinery.

[52] Barzin Doroodgar and Goldie Nejat. A hierarchical reinforcement learning based control architecture for semi-autonomous rescue robots in cluttered environments. In *2010 IEEE International Conference on Automation Science and Engineering*, pages 948–953. IEEE, 2010.

[53] Grazia D'Onofrio, Laura Fiorini, Hiroshi Hoshino, Aiko Matsumori, Yasuo Okabe, Masahiko Tsukamoto, Raffaele Limosani, Alessandra Vitanza, Francesca Greco, Antonio Greco, et al. Assistive robots for socialization in elderly people: results pertaining to the needs of the users. *Aging clinical and experimental research*, 31(9):1313–1329, 2019.

[54] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[55] Paul Ekman. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16, 1999.

[56] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*, 2017.

[57] Federal Emergency Management Agency (FEMA). About the agency.

[58] Yi Feng and Shaoze Cui. A review of emergency response in disasters: Present and future perspectives. *Natural Hazards*, 105(1):1109–1138, 2021.

[59] Thomas Fincannon, Laura Elizabeth Barnes, Robin R Murphy, and Dawn L Riddle. Evidence of the need for social intelligence in rescue robots. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 2, pages 1089–1095. IEEE, 2004.

[60] Alberto Finzi and Andrea Orlandini. Human-robot interaction through mixed-initiative planning for rescue and search rovers. In *Congress of the Italian Association for Artificial Intelligence*, pages 483–494. Springer, 2005.

[61] Avi Fleischer, Alan D Mead, and Jialin Huang. Inattentive responding in mturk and other online samples. *Industrial and Organizational Psychology*, 8(2):196, 2015.

[62] Frederik Funke and Ulf-Dietrich Reips. Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field methods*, 24(3):310–327, 2012.

[63] Márta Gácsi, Anna Kis, Tamás Faragó, Mariusz Janiak, Robert Muszyński, and Ádám Miklósi. Humans attribute emotions to a robot that shows simple behavioural patterns borrowed from dog behaviour. *Computers in Human Behavior*, 59:411–419, 2016.

[64] Moojan Ghafurian, Sami Alperen Akgun, Mark Crowley, and Kerstin Dautenhahn. Recognition of a robot's affective expressions under conditions with limited visibility. In *IFIP Conference on Human-Computer Interaction (INTERACT)*, 2021.

[65] Moojan Ghafurian, Neil Budnarain, and Jesse Hoey. Role of emotions in perception of humanness of virtual agents. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19, page 1979–1981, Richland, SC, 2019. International Foundation for Autonomous Agents and Multiagent Systems.

[66] Moojan Ghafurian, Gabriella Lakatos, Zhuofu Tao, and Kerstin Dautenhahn. Design and evaluation of affective expressions of a zoomorphic robot. In *International Conference on Social Robotics*, pages 1–12. Springer, 2020.

[67] Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. Detecting emotion stimuli in emotion-bearing sentences. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 152–165. Springer, 2015.

[68] Angela Giambattista, Luís Teixeira, Hande Ayanoğlu, Magda Saraiva, and Emília Duarte. Expression of emotions by a service robot: a pilot study. In *International Conference of Design, User Experience, and Usability*, pages 328–336. Springer, 2016.

[69] C. S. Goodman and D. E. Hogan. Urban search and rescue. In *Disaster Medicine*, volume 2, pages 72–85, 2007.

[70] Joseph K Goodman, Cynthia E Cryder, and Amar Cheema. Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, 26(3):213–224, 2013.

[71] Michael A Goodrich, Bryan S Morse, Damon Gerhardt, Joseph L Cooper, Morgan Quigley, Julie A Adams, and Curtis Humphrey. Supporting wilderness search and rescue using a camera-equipped mini uav. *Journal of Field Robotics*, 25(1-2):89–110, 2008.

[72] Ian Greatbatch, Rebecca J Gosling, and Sophie Allen. Quantifying search dog effectiveness in a terrestrial search and rescue environment. *Wilderness & Environmental Medicine*, 26(3):327–334, 2015.

[73] Sharon Gwaltney-Brant, Lisa Murphy, Tina Wismer, and Jay Albretsen. General toxicologic hazards and risks for search-and-rescue dogs responding to urban disasters. *Journal of the American Veterinary Medical Association*, 222:292–5, 03 2003.

[74] Y. Hada and O. Takizawa. 2-4 development of communication technology for search and rescue robots. *Journal of the National Institute of Information and Communications Technology*, 58:131–151, 03 2011.

[75] Quirin Hamp, Leonhard Reindl, and Denise Güthlin. Decision-making behaviour during urban search and rescue: a case study of germany. *Disasters*, 38(1):84–107, 2014.

[76] Victoria Hampshire and Sheilah Robertson. Using the facial grimace scale to evaluate rabbit wellness in post-procedural monitoring. *Lab animal*, 44(7):259, 2015.

[77] Masashi Hasegawa, Nobuyo Ohtani, and Mitsuaki Ohta. Dogs' body language relevant to learning achievement. *Animals*, 4(1):45–58, 2014.

[78] David R Heise. *Expressive order: Confirming sentiments in social actions*. Springer Science & Business Media, 2007.

[79] Jorunn L Helbostad, Beatrix Vereijken, Karin Hesseberg, and Olav Sletvold. Altered vision destabilizes gait in older persons. *Gait & posture*, 30(2):233–238, 2009.

[80] Thomas Hellström and Suna Bensch. Understandable robots. *Paladyn, Journal of Behavioral Robotics*, 9:110–123, 07 2018.

[81] Laura Hoffmann, Melanie Derksen, and Stefan Kopp. What a pity, pepper! how warmth in robots' language impacts reactions to errors during a collaborative task. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '20, page 245–247, New York, NY, USA, 2020. Association for Computing Machinery.

[82] D.E. Hogan and J.L. Burstein. *Disaster Medicine*. Lippincott Williams & Wilkins, 2007.

[83] Ruud Hortensius, Felix Hekele, and Emily S Cross. The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):852–864, 2018.

[84] Eric K Hung and David A Townes. Search and rescue in yosemite national park: a 10-year review. *Wilderness & environmental medicine*, 18(2):111–116, 2007.

[85] Matthew Hutson. Searching for survivors of the mexico earthquake—with snake robots. *Science*, 2017.

[86] Md Jahidul Islam, Marc Ho, and Junaed Sattar. Understanding human motion and gestures for underwater human–robot collaboration. *Journal of Field Robotics*, 36(5):851–873, 2019.

[87] Serkan İzmirli. Considering social presence in the designing of ubiquitous learning environments. In *Managing and Designing Online Courses in Ubiquitous Learning Environments*, pages 201–214. IGI Global, 2020.

[88] Péter Jackovics. Standard of operation for cave rescue in hungary. *International Fire Fighter*, 2016(9):84–86, 2016.

[89] William James. *What is an Emotion?* Simon and Schuster, 2013.

[90] Xiang Ji, Hu Liu, Zizhao Wang, Yuge Li, Yongliang Tian, and Mingyu Huo. Agent-based mr simulation as a tool for combat search and rescue command and control. In *Proceedings of the 2019 3rd International Conference on Artificial Intelligence and Virtual Reality*, pages 1–7, 2019.

[91] S. M. Jimenez Zafra, M. T. Martin Valdivia, E. Martinez Camara, and L. A. Urena Lopez. Studying the scope of negation for spanish sentiment analysis on twitter. *IEEE Transactions on Affective Computing*, 10(1):129–141, 2019.

[92] Brennan Jones, Anthony Tang, and Carman Neustaedter. Remote communication in wilderness search and rescue: Implications for the design of emergency distributed-collaboration tools for network-sparse environments. *Proc. ACM Hum.-Comput. Interact.*, 4(GROUP), January 2020.

[93] Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Nathan G Freier, Rachel L Severson, Brian T Gill, Jolina H Ruckert, and Solace Shen. "robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology*, 48(2):303, 2012.

[94] Ichi Kanaya, Meina Tawaki, and Keiko Yamamoto. Cross-cultural design of facial expressions of robots. In *International Conference on Human-Computer Interaction*, pages 591–605. Springer, 2020.

[95] Yunus Karaca, Mustafa Cicek, Ozgur Tatli, Aynur Sahin, Sinan Pasli, Muhammed Fatih Beser, and Suleyman Turedi. The potential use of unmanned aircraft systems (drones) in mountain search and rescue operations. *The American journal of emergency medicine*, 36(4):583–588, 2018.

[96] M. Karg, A. Samadani, R. Gorbet, K. Kühnlenz, J. Hoey, and D. Kulić. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing*, 4(4):341–359, 2013.

[97] S Karma, E Zorba, GC Pallis, G Statheropoulos, I Balta, K Mikedi, J Vamvakari, A Pappa, M Chalaris, G Xanthopoulos, et al. Use of unmanned vehicles in search and rescue operations in forest fires: Advantages and limitations observed in a field trial. *International journal of disaster risk reduction*, 13:307–312, 2015.

[98] Yuki Kayukawa, Yasutake Takahashi, Takuya Tsujimoto, Kazunori Terada, and Hiroyuki Inoue. Influence of emotional expression of real humanoid robot to human decision-making. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2017.

[99] Stephanie CJ Keating, Aurelie A Thomas, Paul A Flecknell, and Matthew C Leach. Evaluation of emla cream for preventing pain during tattooing of rabbits: changes in physiological, behavioural and facial expression responses. *PloS one*, 7(9), 2012.

[100] Rajiv Khosla, Mei-Tai Chu, and Khanh Nguyen. Affective robot enabled capacity and quality improvement of nursing home aged care services in australia. In *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*, pages 409–414. IEEE, 2013.

[101] Jong-Hwan Kim, Joseph W Starr, and Brian Y Lattimer. Firefighting robot stereo infrared vision and radar sensor fusion for imaging through smoke. *Fire Technology*, 51(4):823–845, 2015.

[102] Takuya Kitade, Satoru Satake, Takayuki Kanda, and Michita Imai. Understanding suitable locations for waiting. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 57–64. IEEE, 2013.

[103] Aniket Kittur, Ed H. Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 453–456, New York, NY, USA, 2008. Association for Computing Machinery.

[104] Barbara Klausz, Anna Kis, Eszter Persa, Ádám Miklósi, and Márta Gácsi. A quick assessment tool for human-directed aggression in pet dogs. *Aggressive behavior*, 40(2):178–188, 2014.

[105] Alexander Kleiner, Christian Dornhege, and Sun Dali. Mapping disaster areas jointly: Rfid-coordinated slam by hurnans and robots. In *2007 IEEE International Workshop on Safety, Security and Rescue Robotics*, pages 1–6. IEEE, 2007.

[106] Andreas Kolling, Phillip Walker, Nilanjan Chakraborty, Katia Sycara, and Michael Lewis. Human Interaction with Robot Swarms: A Survey. *IEEE Transactions on Human-Machine Systems*, 46(1):9–26, 2016.

[107] Veronika Konok, Antal Dóka, and Ádám Miklósi. The behavior of the domestic dog (canis familiaris) during separation from and reunion with the owner: A questionnaire and an experimental study. *Applied Animal Behaviour Science*, 135(4):300–308, 2011.

[108] Veronika Konok, Krisztina Nagy, and Ádám Miklósi. How do humans represent the emotions of dogs? the resemblance between the human representation of the canine and the human affective space. *Applied Animal Behaviour Science*, 162:37–46, 2015.

[109] Beáta Korcsok, Veronika Konok, György Persa, Tamás Faragó, Mihoko Niitsuma, Ádám Miklósi, Péter Korondi, Péter Baranyi, and Márta Gácsi. Biologically inspired emotional expressions for artificial agents. *Frontiers in psychology*, 9:1191, 2018.

[110] Austin Kothig, John Munoz, Sami Alperen Akgun, Alexander M. Aroyo, and Kerstin Dautenhahn. Connecting humans and robots using physiological signals – closing-the-loop in hri. In *To appear in 2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, page 0. IEEE, 2021.

[111] Geert-Jan M Kruijff, M Janíček, Shanker Keshavdas, Benoit Larochelle, Hendrik Zender, Nanja JJM Smets, Tina Mioch, Mark A Neerincx, JV Diggelen, Francis Colas, et al. Experience in system design for human-robot teaming in urban search and rescue. In *Field and Service Robotics*, pages 111–125. Springer, 2014.

[112] Gabriella Lakatos, Márta Gácsi, Veronika Konok, Ildiko Bruder, Boróka Bereczky, Peter Korondi, and Adam Miklosi. Emotion attribution to a non-humanoid robot in different social situations. *PloS one*, 9(12):e114207, 2014.

[113] Benoit Larochelle, Geert-Jan M Kruijff, Nanja Smets, Tina Mioch, and Peter Groenewegen. *Establishing human situation awareness using a multi-modal operator control unit in an urban search & rescue human-robot team*. IEEE, 2011.

[114] Joseph E LeDoux. Chapter 21 - Evolution of human emotion: A view through fear. In Michel A Hofman and Dean Falk, editors, *Evolution of the Primate Brain*, volume 195 of *Progress in Brain Research*, pages 431–442. Elsevier, 2012.

[115] Heather Lench, Sarah Flores, and Shane Bench. Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological bulletin*, 137:834–55, 09 2011.

[116] Michael Lewis, Huadong Wang, Shih-Yi Chien, Paul Scerri, Prasanna Velagapudi, Katia Sycara, and Breelyn Kane. Teams organization and performance in multi-human/multi-robot teams. In *2010 IEEE International Conference on Systems, Man and Cybernetics*, pages 1617–1623. IEEE, 2010.

[117] Jamy Li, Daniel Davison, Bob Schadenberg, Pauline Chevalier, Alyssa Alcorn, Alria Williams, Suncica Petrovic, Snezana Babovic Dimitrijevic, Jie Shen, Liz Pellicano, et al. Usability of a robot's realistic facial expressions and peripherals in autistic children's therapy. *arXiv preprint arXiv:2007.12236*, 2020.

[118] T. Linder, V. Tretyakov, S. Blumenthal, P. Molitor, D. Holz, R. Murphy, S. Tadokoro, and H. Surmann. Rescue robots at the collapse of the municipal archive of cologne city: A field report. In *2010 IEEE Safety Security and Rescue Robotics*, pages 1–6, July 2010.

[119] Kristen Lindquist, Erika Siegel, Karen Quigley, and Lisa Barrett. The hundred-year emotion war: Are emotions natural kinds or psychological constructions? comment on lench, flores, and bench (2011). *Psychological bulletin*, 139:255–63, 01 2013.

[120] Jinguo Liu, Yuechao Wang, Bin Li, and Shugen Ma. Current research, key performances and future development of search and rescue robot. *Jixie Gongcheng Xuebao/Chinese Journal of Mechanical Engineering*, 42(12):1–12, 2006.

[121] Yugang Liu and Goldie Nejat. Robotic urban search and rescue: A survey from the control perspective. *Journal of Intelligent & Robotic Systems*, 72(2):147–165, 2013.

[122] Eleftherios Lygouras, Nicholas Santavas, Anastasios Taitzoglou, Konstantinos Tarchanidis, Athanasios Mitropoulos, and Antonios Gasteratos. Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations. *Sensors*, 19(16):3542, 2019.

[123] Evgeni Magid, Artyom Pashkin, Nikita Simakov, Bulat Abbyasov, Jackrit Suthakorn, Mikhail Svinin, and Fumitoshi Matsuno. Artificial intelligence based framework for robotic search and rescue operations conducted jointly by international teams. In *Proceedings of 14th International Conference on Electromechanics and Robotics "Zavalishin's Readings"*, pages 15–26. Springer, 2020.

[124] Hamza Mahdi, Sami Alperen Akgun, Shahed Saleh, and Kerstin Dautenhahn. A survey on design and evolution of social robots — past, present and future. *Submitted to Robotics and Autonomous Systems*, 0:0, 2021.

[125] Marco Manca, Fabio Paternò, Carmen Santoro, Eleonora Zedda, Chiara Braschi, Roberta Franco, and Alessandro Sale. The impact of serious games with humanoid robots on mild cognitive impairment older adults. *International Journal of Human-Computer Studies*, page 102509, 2020.

[126] John Marx, Ron Walls, and Robert Hockberger. *Rosen's emergency medicine-concepts and clinical practice*. Elsevier Health Sciences, 2013.

[127] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. The effect of visual appearance on the performance of continuous sliders and visual analogue scales. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5421–5432, 2016.

[128] Aníbal Matos, Alfredo Martins, André Dias, Bruno Ferreira, José Miguel Almeida, Hugo Ferreira, Guilherme Amaral, André Figueiredo, Rui Almeida, and Filipe Silva.

Multiple robot operations for maritime search and rescue in eurathlon 2015 competition. In *OCEANS 2016-Shanghai*, pages 1–7. IEEE, 2016.

[129] Lynn C Matsumiya, Robert E Sorge, Susana G Sotocinal, John M Tabaka, Jeffrey S Wieskopf, Austin Zaloum, Oliver D King, and Jeffrey S Mogil. Using the mouse grimace scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. *Journal of the american association for laboratory animal science*, 51(1):42–49, 2012.

[130] Fumitoshi Matsuno, Noritaka Sato, Kazuyuki Kon, Hiroki Igarashi, Tetsuya Kimura, and Robin Murphy. Utilization of robot systems in disaster sites of the great eastern japan earthquake. In *Field and service robotics*, pages 1–17. Springer, 2014.

[131] Sven Mayer, Lars Lischke, and Pawel W. Woźniak. Drones for Search and Rescue. In *1st International Workshop on Human-Drone Interaction*, Glasgow, United Kingdom, May 2019. Ecole Nationale de l'Aviation Civile [ENAC].

[132] Conor McGinn, Eamonn Bourke, Andrew Murtagh, Cian Donovan, Patrick Lynch, Michael F Cullinan, and Kevin Kelly. Meet stevie: a socially assistive robot developed through application of a 'design-thinking'approach. *Journal of Intelligent & Robotic Systems*, 98(1):39–58, 2020.

[133] DM McGuigan, BL Deam, and DK Bull. Urban search and rescue and the role of the engineer. *Unpublished master's thesis, University of Canterbury, New Zealand*, 2002.

[134] Albert Mehrabian. Framework for a comprehensive description and measurement of emotional states. *Genetic, social, and general psychology monographs*, 1995.

[135] Nathan Michael, Shaojie Shen, Kartik Mohta, Vijay Kumar, Keiji Nagatani, Yoshito Okada, Seiga Kiribayashi, Kazuki Otake, Kazuya Yoshida, Kazunori Ohno, et al. Collaborative mapping of an earthquake damaged building via ground and aerial robots. In *Field and service robotics*, pages 33–47. Springer, 2014.

[136] Ádám Miklósi, Péter Korondi, Vicente Matellán, and Márta Gácsi. Ethorobotics: A new approach to human-robot relationship. *Frontiers in psychology*, 8:958, 2017.

[137] GS Miller. 13 snake robots for search and rescue. *Neurotechnology for Biomimetic Robots, MIT Press, Cambridge, MA*, page 271, 2002.

[138] Nidhi Mishra, Manoj Ramanathan, Ranjan Satapathy, Erik Cambria, and Nadia Magnenat-Thalmann. Can a humanoid robot be part of the organizational workforce? a user study leveraging sentiment analysis. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7. IEEE, 2019.

[139] Ben Mitchinson and Tony J Prescott. Miro: a robot "mammal" with a biomimetic brain-based control system. In *Conference on Biomimetic and Biohybrid Systems*, pages 179–191. Springer, 2016.

[140] Paul H Morris, Christine Doe, and Emma Godsell. Secondary emotions in non-primate species? behavioural reports and subjective claims by animal owners. *Cognition and emotion*, 22(1):3–20, 2008.

[141] J.P. Moss. *North Central Texas Search And Rescue Operational/By-Laws Manual Dedication to Frank C. Clark C.E.O./ Col.* Lulu.com, 2013.

[142] Anastasios I Mourikis, Nikolas Trawny, Stergios I Roumeliotis, Daniel M Helmick, and Larry Matthies. Autonomous stair climbing for tracked vehicles. *The International Journal of Robotics Research*, 26(7):737–758, 2007.

[143] Lin Mu and Enjin Zhao. The optimization of maritime search and rescue simulation system based on cps. In *Big Data Analytics for Cyber-Physical Systems*, pages 231–245. Springer, 2020.

[144] R. R. Murphy. Trial by fire [rescue robots]. *IEEE Robotics Automation Magazine*, 11(3):50–61, 2004.

[145] Robin R Murphy. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(2):138–153, 2004.

[146] Robin R Murphy. Trial by fire [rescue robots]. *IEEE Robotics & Automation Magazine*, 11(3):50–61, 2004.

[147] Robin R Murphy. *Disaster robotics.* MIT press, 2014.

[148] Robin R Murphy, Dawn Riddle, and Eric Rasmussen. Robot-assisted medical reachback: a survey of how medical personnel expect to interact with rescue robots. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*, pages 301–306. IEEE, 2004.

[149] Dula Nad, Christopher Walker, Igor Kvasić, Derek Orbaugh Antillon, Nikola Mišković, Iain Anderson, and Ivan Lončar. Towards advancing diver-robot interaction capabilities. *IFAC-PapersOnLine*, 52(21):199–204, 2019.

[150] Ambreen Nazir, Yuan Rao, Lianwei Wu, and Ling Sun. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, 2020.

[151] Rock Networks. Gps enabled two way radios make difference with search teams.

[152] Niels A Nijdam. Mapping emotion to color. *Book Mapping emotion to color*, pages 2–9, 2009.

[153] Farzad Niroui, Kaicheng Zhang, Zendai Kashino, and Goldie Nejat. Deep reinforcement learning robot for search and rescue applications: Exploration in unknown cluttered environments. *IEEE Robotics and Automation Letters*, 4(2):610–617, 2019.

[154] T. Nomura and K. Kawakami. Relationships between robot's self-disclosures and human's anxiety toward robots. In *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, volume 3, pages 66–69, Aug 2011.

[155] Heather L O'Brien and Elaine G Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6):938–955, 2008.

[156] Justice Institute of British Colombia. Search and rescue management level 1 participant manual, 2015.

[157] U.s. Department of Homeland Security, F.E.M. Agency, and U.S.F. Administration. *Firefighter Fatalities in the United States in 1995*. CreateSpace Independent Publishing Platform, 2013.

[158] Yosuke Okita, Manabu Sugita, Tsukasa Katsube, and Yusuke Minato. Capacity building of international search and rescue teams through the classification system: Example from japan disaster relief team and insarag external classification and reclassification. *Journal of Japan Association for Earthquake Engineering*, 18(6):6_23–6_40, 2018.

[159] Richard S Olson and Robert A Olson. Urban heavy rescue. *Earthquake Spectra*, 3(4):645–658, 1987.

[160] Yang Pan, Feng Gao, Chenkun Qi, and Xun Chai. Human-tracking strategies for a six-legged rescue robot based on distance and view. *Chinese Journal of Mechanical Engineering*, 29(2):219–230, 2016.

[161] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. Running experiments on Amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419, 2010.

[162] Srinivas Parthasarathy and Carlos Busso. Jointly predicting arousal, valence and dominance with multi-task learning. In *Interspeech*, pages 1103–1107, 2017.

[163] Isabella Poggi, Catherine Pelachaud, Fiorella de Rosis, Valeria Carofiglio, and Berardina De Carolis. Greta. a believable embodied conversational agent. In *Multimodal intelligent information presentation*, pages 3–25. Springer, 2005.

[164] Péter Pongrácz, Csaba Molnár, Antal Dóka, and Ádám Miklósi. Do children understand man's best friend? classification of dog barks by pre-adolescents and adults. *Applied animal behaviour science*, 135(1-2):95–102, 2011.

[165] Péter Pongrácz, Csaba Molnár, Adám Miklósi, and Vilmos Csányi. Human listeners are able to classify dog (canis familiaris) barks recorded in different situations. *Journal of comparative psychology*, 119(2):136, 2005.

[166] K Pratt, RR Murphy, S Stover, and C Griffin. Requirements for semi-autonomous flight in miniature uavs for structural inspection. *AUVSI's Unmanned Systems North America. Orlando, Florida, Association for Unmanned Vehicle Systems International*, 2006.

[167] R Cynthia Monica Priya and JGR Sathiaseelan. An explorative study on sentiment analysis. In *2017 World Congress on Computing and Communication Technologies (WCCCT)*, pages 140–142. IEEE, 2017.

[168] A Quaranta, M Siniscalchi, and G Vallortigara. Asymmetric tail-wagging responses by dogs to different emotive stimuli. *Current Biology*, 17(6):R199–R201, 2007.

[169] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Ng. Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, volume 3, 01 2009.

[170] KJ Rafferty and Euan William McGookin. An autonomous air-sea rescue system using particle swarm optimization. In *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, pages 459–464. IEEE, 2013.

[171] Daniel J Rea, James E Young, and Pourang Irani. The roomba mood ring: an ambient-display robot. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 217–218, 2012.

[172] Rainer Reisenzein, Sandra Bördgen, Thomas Holtbernd, and Denise Matz. Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of personality and social psychology*, 91(2):295, 2006.

[173] Paula V. Resnik and H. Bruce Lammers. The influence of self-esteem on cognitive responses to machine-like versus human-like computer feedback. *The Journal of Social Psychology*, 125(6):761–769, 1985.

[174] H. Ritschel, I. Aslan, S. Mertes, A. Seiderer, and E. André. Personalized synthesis of intentional and emotional non-verbal sounds for social robots. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–7, 2019.

[175] AJA Rivera, ADC Villalobos, JCN Monje, JAG Mariñas, and CM Oppus. Post-disaster rescue facility: Human detection and geolocation using aerial drones. In *2016 IEEE Region 10 Conference (TENCON)*, pages 384–386. IEEE, 2016.

[176] Elise WB Russell. *Real-time topic and sentiment analysis in human-robot conversation*. PhD thesis, Marquette University, 2015.

[177] Alessandra Maria Sabelli, Takayuki Kanda, and Norihiro Hagita. A conversational robot in an elderly care center: an ethnographic study. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 37–44. IEEE, 2011.

[178] Jelle Saldien, Kristof Goris, Bram Vanderborght, Johan Vanderfaeillie, and Dirk Lefeber. Expressing emotions with the social robot probo. *International Journal of Social Robotics*, 2(4):377–389, 2010.

[179] Jorge Sales, Raul Marin, Enric Cervera, Sergio Rodríguez, and Javier Pérez. Multi-sensor person following in low-visibility scenarios. *Sensors*, 10(12):10953–10966, 2010.

[180] Magda Saraiva, Hande Ayanoğlu, and Beste Özcan. Emotional design and human-robot interaction. In *Emotional Design in Human-Robot Interaction*, pages 119–141. Springer, 2019.

[181] Noritaka Sato, Fumitoshi Matsuno, Tatsuhiro Yamasaki, Tetsushi Kamegawa, Naoji Shiroma, and Hiroki Igarashi. Cooperative task execution by a multiple robot team and its operators in search and rescue operations. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 2, pages 1083–1088. IEEE, 2004.

[182] Annett Schirmer, Cui Shan Seow, and Trevor B Penney. Humans process dog and human facial affect in similar ways. *PLoS One*, 8(9), 2013.

[183] S Scuri, F Petrelli, TTC Nguyen, and I Grappasonni. Training to improve resilience and coping to monitor ptsd in rescue workers. *Journal of preventive medicine and hygiene*, 60(1):E58, 2019.

[184] Y. S. Sefidgar, K. E. MacLean, S. Yohanan, H. F. M. Van der Loos, E. A. Croft, and E. J. Garland. Design and evaluation of a touch-centered calming interaction with a social robot. *IEEE Transactions on Affective Computing*, 7(2):108–121, 2016.

[185] Hannes Seppänen, Jaana Mäkelä, Pekka Luokkala, and Kirsi Virrantaus. Developing shared situational awareness for emergency management. *Safety science*, 55:1–9, 2013.

[186] R. Setchi and O. K. Asikhia. Exploring user experience with image schemas, sentiments, and semantics. *IEEE Transactions on Affective Computing*, 10(2):182–195, 2019.

[187] Megha Sharma, Dale Hildebrandt, Gem Newman, James E Young, and Rasit Eskicioglu. Communicating affect via flight path exploring use of the laban effort system for designing affective locomotion paths. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 293–300. IEEE, 2013.

[188] Davood Shiri, Vahid Akbari, and F Sibel Salman. Online routing and scheduling of search-and-rescue teams. *OR Spectrum*, pages 1–30, 2020.

[189] Mario Silvagni, Andrea Tonoli, Enrico Zenerino, and Marcello Chiaberge. Multipurpose uav for search and rescue operations in mountain avalanche events. *Geomatics, Natural Hazards and Risk*, 8(1):18–33, 2017.

[190] Kimberly A Slensky, Kenneth J Drobatz, Amanda B Downend, and Cynthia M Otto. Deployment morbidity among search-and-rescue dogs used after the september 11, 2001, terrorist attacks. *Journal of the American Veterinary Medical Association*, 225(6):868–873, 2004.

[191] Lynn Smith-Lovin, Dawn T Robinson, Bryan C Cannon, Jesse K Clark, Robert Freeland, Jonathan H Morgan, and Kimberly B Rogers. Mean affective ratings of 929 identities, 814 behaviors, and 660 modifiers by university of georgia and duke university undergraduates and by community members in durham, nc, in 2012-2014. *University of Georgia: Distributed at UGA Affect Control Theory Website: http://research. franklin. uga. edu/act*, 2016.

[192] Mads Soegaard and Rikke Friis Dam. The encyclopedia of human-computer interaction. *The encyclopedia of human-computer interaction*, 2012.

[193] Sichao Song and Seiji Yamada. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 2–11. IEEE, 2017.

[194] Sichao Song and Seiji Yamada. Designing led lights for communicating gaze with appearance-constrained robots. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 94–97. IEEE, 2018.

[195] Stefan Sosnowski, Ansgar Bittermann, Kolja Kuhnlenz, and Martin Buss. Design and evaluation of emotion-display eddie. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3113–3118. IEEE, 2006.

[196] Joseph W Starr and BY Lattimer. Evaluation of navigation sensors in fire smoke environments. *Fire Technology*, 50(6):1459–1481, 2014.

[197] Anna Stenzel, Eris Chinellato, Maria A Tirado Bou, Ángel P del Pobil, Markus Lappe, and Roman Liepelt. When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance*, 38(5):1073, 2012.

[198] Sandy Stott. *Critical Hours: Search and Rescue in the White Mountains*. University Press of New England, 2018.

[199] Gabriella Tami and Anne Gallagher. Description of the behaviour of domestic dog (canis familiaris) by experienced and inexperienced people. *Applied Animal Behaviour Science*, 120(3-4):159–169, 2009.

[200] Mark Meerum Terwogt and Jan B Hoeksma. Colors and emotions: Preferences and combinations. *The Journal of General Psychology*, 122(1):5–17, 1995.

[201] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

[202] József Topál, Ádám Miklósi, Vilmos Csányi, and Antal Dóka. Attachment behavior in dogs (canis familiaris): a new application of ainsworth's (1969) strange situation test. *Journal of comparative psychology*, 112(3):219, 1998.

[203] Horst Treiblmaier and P. Filzmoser. Benefits from using continuous rating scales in online survey research. In *ICIS*, 2011.

[204] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[205] Bram Van Acker, Davy Parmentier, Peter Vlerick, and Jelle Saldien. Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, Technology & Work*, 04 2018.

[206] Christopher Van Tilburg, Colin K Grissom, Ken Zafren, Scott McIntosh, Martin I Radwin, Peter Paal, Pascal Haegeli, Albert R Wheeler, David Weber, Bruce Tremper, et al. Wilderness medical society practice guidelines for prevention and management of avalanche and nonavalanche snow burial accidents. *Wilderness & environmental medicine*, 28(1):23–42, 2017.

[207] Judit Vas, József Topál, Márta Gácsi, Adám Miklósi, and Vilmos Csányi. A friend or an enemy? dogs' reaction to an unfamiliar person showing behavioural cues of threat and friendliness at different times. *Applied Animal Behaviour Science*, 94(1-2):99–115, 2005.

[208] Juan David Velásquez. An emotion-based approach to robotics. In *Proc. International Conference on Intelligent Robots and Systems.*, volume 1, pages 235–240. IEEE, 1999.

[209] Sebastian Wallkotter, Rebecca Stower, Arvid Kappas, and Ginevra Castellano. A robot by any other frame: Framing and behaviour influence mind perception in virtual but not real-world environments. *ACM/IEEE International Conference on Human-Robot Interaction*, pages 609–618, 2020.

[210] Adam Waytz, John Cacioppo, and Nicholas Epley. Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3):219–232, 2010.

[211] Ryan Wegner and John Anderson. Agent-based support for balancing teleoperation and autonomy in urban search and rescue. *International Journal of Robotics and Automation*, 21(2):120–128, 2006.

[212] Yair Weiss, Eero P Simoncelli, and Edward H Adelson. Motion illusions as optimal percepts. *Nature neuroscience*, 5(6):598–604, 2002.

[213] Sherri C Widen and James A Russell. The "disgust face" conveys anger to children. *Emotion*, 10(4):455, 2010.

[214] Sherri C Widen, James A Russell, and Aimee Brooks. Anger and disgust: Discrete or overlapping categories. In *2004 APS Annual Convention, Boston College, Chicago, IL*, 2004.

[215] Arnold Wilkins, Jennifer Veitch, and Brad Lehman. Led lighting flicker and potential health concerns: Ieee standard par1789 update. In *2010 IEEE Energy Conversion Congress and Exposition*, pages 171–178. IEEE, 2010.

[216] Sarah N Woods, Michael L Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Methodological issues in hri: A comparison of live and video-based methods in robot to human approach direction trials. In *ROMAN 2006-the 15th IEEE international symposium on robot and human interactive communication*, pages 51–58. IEEE, 2006.

[217] Changlong Ye, Shugen Ma, and Bin Li. Design and basic experiments of a shape-shifting mobile robot for urban search and rescue. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3994–3999. IEEE, 2006.

[218] Jingchao Zhao, Junyao Gao, Fangzhou Zhao, and Yi Liu. A search-and-rescue robot system for remotely sensing the underground coal mine environment. *Sensors*, 17(10):2426, 2017.

[219] Xiqian Zheng, Masahiro Shiomi, Takashi Minato, and Hiroshi Ishiguro. What kinds of robot's touch will match expressed emotions? *IEEE Robotics and Automation Letters*, 5(1):127–134, 2019.

APPENDICES

Appendix A

Ethics Clearance Certificates

UNIVERSITY OF WATERLOO

Notification of Ethics Clearance to Conduct Research with Human Participants

---

Principal Investigator: Kerstin Dautenhahn (Electrical and Computer Engineering)
Student investigator: Sami Alperen Akgun (Systems Design Engineering)
Collaborator: Moojan Ghafurian (Computer Science Computing Facility)
File #: 41900
Title: How Robots' Actions Map to Emotions in an Urban Search and Rescue Situation

---

The Human Research Ethics Committee is pleased to inform you this study has been reviewed and given ethics clearance.

Initial Approval Date: 04/07/20 (m/d/y)

University of Waterloo Research Ethics Committees are composed in accordance with, and carry out their functions and operate in a manner consistent with, the institution's guidelines for research with human participants, the Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans (TCPS, 2nd edition), International Conference on Harmonization: Good Clinical Practice (ICH-GCP), the Ontario Personal Health Information Protection Act (PHIPA), the applicable laws and regulations of the province of Ontario. Both Committees are registered with the U.S. Department of Health and Human Services under the Federal Wide Assurance, FWA00021410, and IRB registration number IRB00002419 (HREC) and IRB00007409 (CREC).

This study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Expiry Date: 04/06/21 (m/d/y)

Multi-year research must be renewed at least once every 12 months unless a more frequent review has otherwise been specified. Studies will only be renewed if the renewal report is received and approved before the expiry date. Failure to submit renewal reports will result in the investigators being notified ethics clearance has been suspended and Research Finance being notified the ethics clearance is no longer valid.

Level of review: Delegated Review

Signed on behalf of the Human Research Ethics Committee

Jo cenaiko

Jo Cenaiko, Ethics Advisor, jrcenaik@uwaterloo.ca, 519-888-4567, ext. 30321

This above named study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Documents reviewed and received ethics clearance for use in the study and/or received for information:

file: end_note_v3.docx
file: Consent_Formv4.docx
file: robot_images.zip
file: Study Materials_v2.docx
file: Mturk_v3.docx

Approved Protocol Version 4 in Research Ethics System
This is an official document. Retain for your files.
You are responsible for obtaining any additional institutional approvals that might be required to complete this study.

UNIVERSITY OF WATERLOO

Notification of Ethics Clearance to Conduct Research with Human Participants

---

Principal Investigator: Kerstin Dautenhahn (Electrical and Computer Engineering)

Co-Principal Investigator: Moojan Ghafurian (School of Computer Science)

Research assistant: Zhuofu Tao (Computer Science Computing Facility)

File #: 41671

Title: Evaluation of Emotional Displays in Miro Robot

The Human Research Ethics Committee is pleased to inform you this study has been reviewed and given ethics clearance.

Initial Approval Date: 01/14/20 (m/d/y)

University of Waterloo Research Ethics Committees are composed in accordance with, and carry out their functions and operate in a manner consistent with, the institution's guidelines for research with human participants, the Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans (TCPS, 2nd edition), International Conference on Harmonization: Good Clinical Practice (ICH-GCP), the Ontario Personal Health Information Protection Act (PHIPA), the applicable laws and regulations of the province of Ontario. Both Committees are registered with the U.S. Department of Health and Human Services under the Federal Wide Assurance, FWA00021410, and IRB registration number IRB00002419 (HREC) and IRB00007409 (CREC).

This study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Expiry Date: 01/15/21 (m/d/y)

Multi-year research must be renewed at least once every 12 months unless a more frequent review has otherwise been specified. Studies will only be renewed if the renewal report is received and approved before the expiry date. Failure to submit renewal reports will result in the investigators being notified ethics clearance has been suspended and Research Finance being notified the ethics clearance is no longer valid.

Level of review: Delegated Review

Signed on behalf of the Human Research Ethics Committee

Heather Root, Senior Manager, Ethics, heather.root@uwaterloo.ca, 519-888-4567, ext. 30469

This above named study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Documents reviewed and received ethics clearance for use in the study and/or received for information:

file: MechanicalTurkTemplate.docx

file: example3.mov

file: example2.mov

file: example1.mov

file: Method.pdf

file: ConsentForm-V2.docx

file: endMessage.pdf

Approved Protocol Version 2 in Research Ethics System

This is an official document. Retain for your files.

You are responsible for obtaining any additional institutional approvals that might be required to complete this study.

UNIVERSITY OF WATERLOO
Notification of Ethics Clearance to Conduct Research with Human Participants

---

Principal Investigator: Kerstin Dautenhahn (Systems Design Engineering)
Student investigator: Sami Alperen Akgun (Systems Design Engineering)
Co-Principal Investigator: Moojan Ghafurian (Electrical and Computer Engineering)
Co-Investigator: Mark Crowley (Electrical and Computer Engineering)
Student investigator: Hamza Mahdi (Electrical and Computer Engineering)
Research assistant: Shahed Saleh (Mechanical and Mechatronics Engineering)
File #: 43033
Title: Evaluation of Common Situations Occurring During Search and Rescue Missions

---

The Human Research Ethics Committee is pleased to inform you this study has been reviewed and given ethics clearance.

Initial Approval Date: 03/30/21 (m/d/y)

University of Waterloo Research Ethics Committees are composed in accordance with, and carry out their functions and operate in a manner consistent with, the institution's guidelines for research with human participants, the Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans (TCPS, 2nd edition), International Conference on Harmonization: Good Clinical Practice (ICH-GCP), the Ontario Personal Health Information Protection Act (PHIPA), the applicable laws and regulations of the province of Ontario. Both Committees are registered with the U.S. Department of Health and Human Services under the Federal Wide Assurance, FWA00021410, and IRB registration number IRB00002419 (HREC) and IRB00007409 (CREC).

This study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Expiry Date: 03/31/22 (m/d/y)

Multi-year research must be renewed at least once every 12 months unless a more frequent review has otherwise been specified. Studies will only be renewed if the renewal report is received and approved before the expiry date. Failure to submit renewal reports will result in the investigators being notified ethics clearance has been suspended and Research Finance being notified the ethics clearance is no longer valid.

Level of review: Delegated Review

Signed on behalf of the Human Research Ethics Committee

Heather Root, Senior Manager, Ethics, heather.root@uwaterloo.ca, 519-888-4567, ext. 30469

This above named study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Documents reviewed and received ethics clearance for use in the study and/or received for information:

file: Protocols.pdf

file: sar_implementation_emotion_v2.docx
file: Method (Procedure)_v2 (1) (1).docx
file: side_fear.mp4
file: husky-pack.jpg
file: questionnaire.odt
file: 4_withText.avi
file: fear_4-20-21.avi
file: ConsentForm_v3 (1).docx
file: end_note-2_v3 (1).docx
Approved Protocol Version 3 in Research Ethics System
This is an official document. Retain for your files.
You are responsible for obtaining any additional institutional approvals that might be required to complete this study.

Appendix B

Online Interface in Experiment 1

Online interface employed in Experiment 1 is given in this appendix.

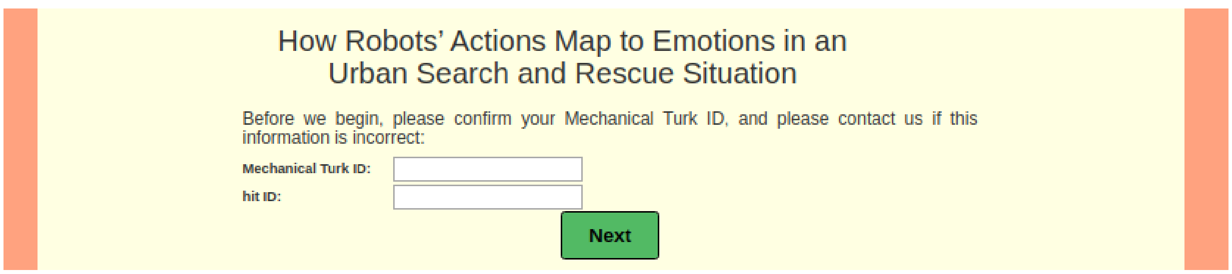


Figure B.1: Login Page

Information and Consent Form

Date: April 2020

Study Title: How Robots' Actions Map to Emotions in an Urban Search and Rescue Situation

We are all from the University of Waterloo in Canada.

Principal Investigator: Prof. Kerstin Dautenhahn, Department of Electrical and Computer Engineering/Systems Design Engineering

Student Investigator: Sami Alperen Akgun, Department of Systems Design Engineering

Collaborator: Dr. Moojan Ghafurian, School of Computer Science

Overview: This is a study to understand how you would map emotions of a rescue robot to possible messages in scenarios that are related to search and rescue situations. We will show you a message and ask you to select emotions that you feel would represent that message. We will also ask you to fill out a questionnaire. Your answer to the questions will be logged. This does not require you to download anything onto your computer.

The outcome of this study will help us to create an alternative and intuitive communication channel between humans and robots during urban search and rescue operations.

Procedure: The procedure of the study is as follows: (1) you will first sign the consent form and (2) read the instructions. (3) you will see an example of a simplified urban search and rescue mission story involving robots and we will show pictures of some existing rescue robots to help you imagine the situation and context. Afterward, you will (4) read about 10 different situations a robot might be in and the corresponding message it wants to send. (1-2 sentences) (5) You will select which emotion(s) should be selected that can represent the messages that the robot wants to convey. Afterwards, (6) you will answer a few sets of questionnaires. In the questionnaires, we will (a) gather demographic information, (b) ask you opinions about rescue robots, (c) ask a few questions about your experience and familiarity with search and rescue applications, as well as emotions in general. The study is expected to take about 25 minutes.

We prefer to receive full responses as they provide valuable information that can help in our study. However, if you do not wish to answer certain questions, you may choose to not answer them. You can do so by selecting "not sure" or "do not wish to answer" if available, or by contacting us at sirrl.waterloo@uwaterloo.ca. **You can complete this HIT only once.**

Duration/Time: The whole process will take up to 25 minutes.

Confidentiality: You will not be identified individually in any written reports of this research. However, your answers in open text form might be used as anonymous quotations. Note that when information is transmitted over the internet, privacy cannot be entirely guaranteed. There is always a risk that your responses may be intercepted by a third party (e.g., government agencies, hackers). Any confidential information (namely your MTurk ID) will be removed from our records within 1 year. You may withdraw your consent and your data within this period by contacting us. The data will be securely stored in locked offices in the research laboratory of Prof. Kerstin Dautenhahn (to which only she has access. Your identity will be kept confidential), and on the principal investigators' computer for analysis. Anonymized data may be published publicly in the future. Anonymized data will be stored for a minimum of 7 years.

Right to Ask Questions: Please contact us at sirrl.waterloo@uwaterloo.ca with questions or concerns about this study. This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE# 41900). If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

Payment for participation: You will be paid \$2 if you complete all the steps of the experiment. If you do not wish to continue at any point, your participation in the study will end and you will be remunerated for the proportion of what you completed (number of questions that you have answered). If you wish to stop at any time, please **do not submit the HIT** and contact us at sirrl.waterloo@uwaterloo.ca. In that case, you will received the amount specified on the HIT plus an additional payment based on the number of questions that you have completed. For example, if you wish to stop in the middle of the study, you will receive (\$1)+ (\$1/2)= \$1.5.

To accept your HIT, you must add the code that is provided to you on the end page, or given to you by us if you wish to end the study earlier. **We will not be able to accept your HIT if you do not provide the correct code.**

Voluntary Participation: Your decision to take part in this research is voluntary. You do not have to answer any questions you do not want to answer. You can stop at any time. If this is the case, please exit the page and abandon the Mechanical Turk HIT. We ask you to contact us in this situation, so we can remunerate you for the proportion of what you completed. In this case your data will be discarded and not be used in the study.

Thank you for your interest in our research and for your assistance with this project. You must be 18 years of age or older to consent to take part in this research study. By accepting this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities. With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.

I Agree

Figure B.2: Information and Consent Form

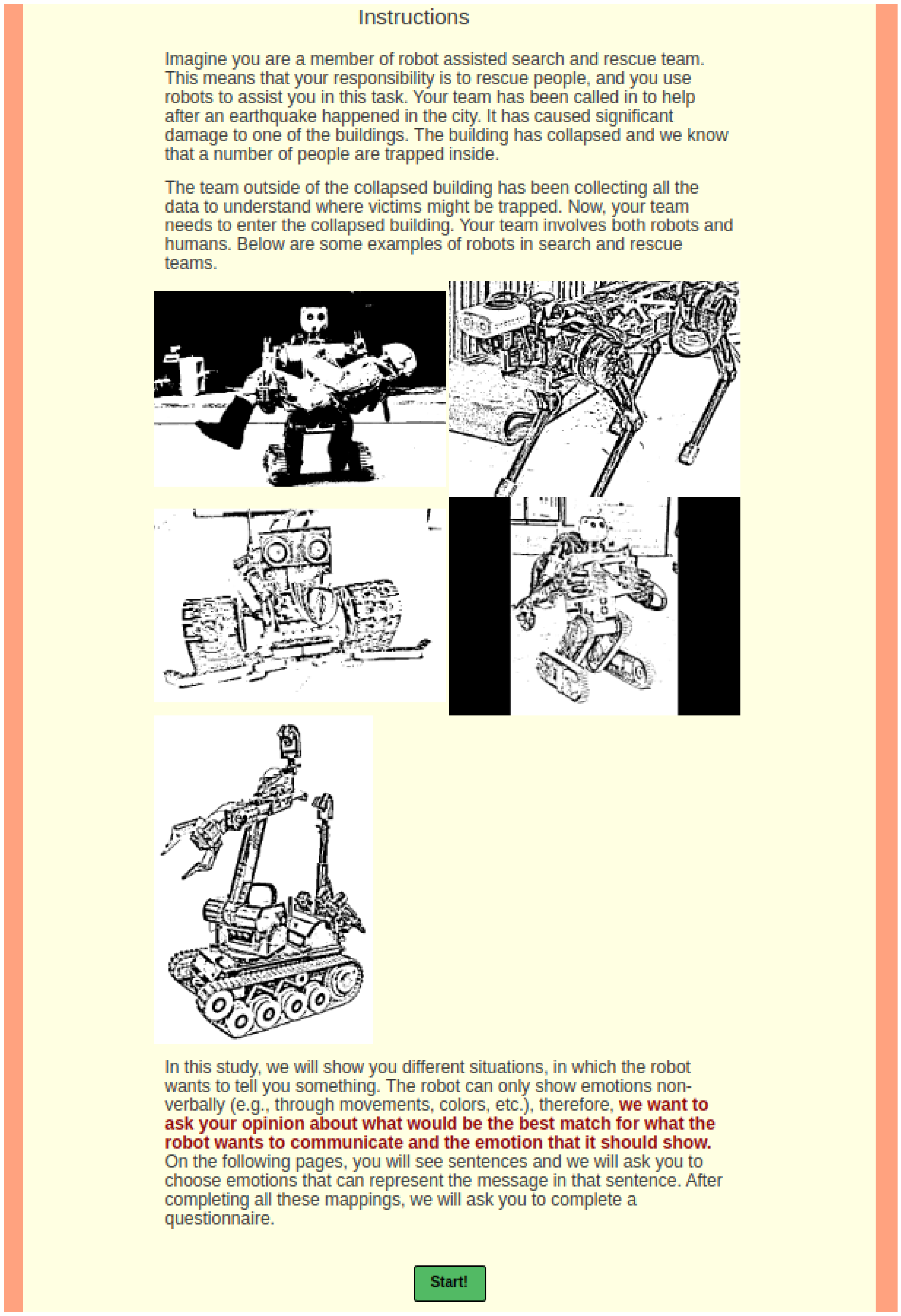


Figure B.3: Instructions

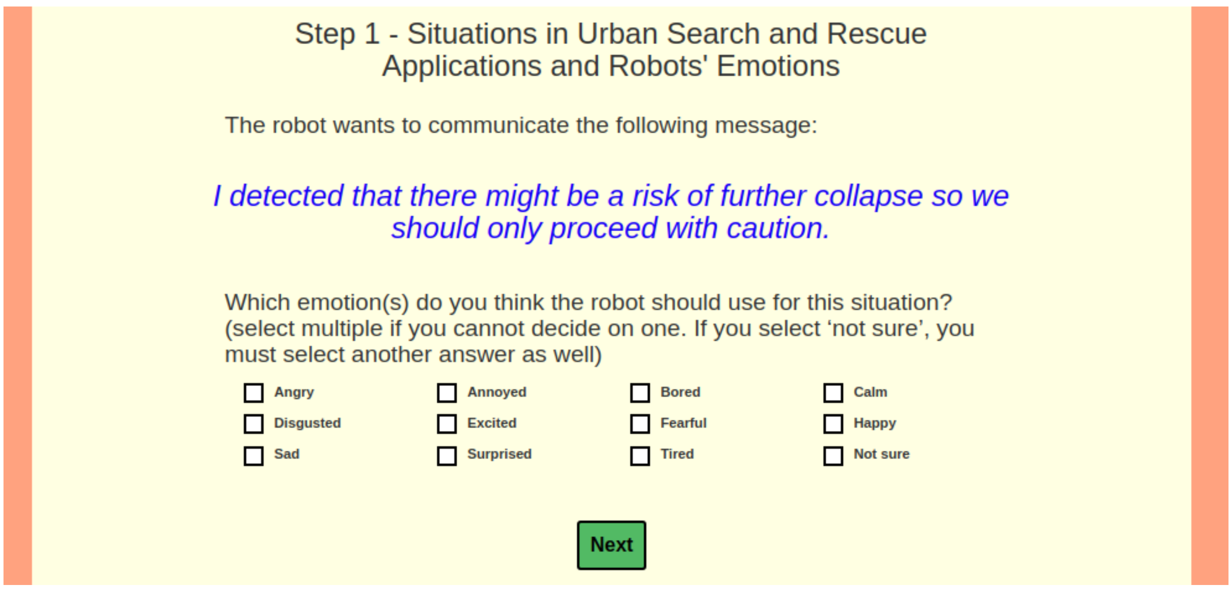
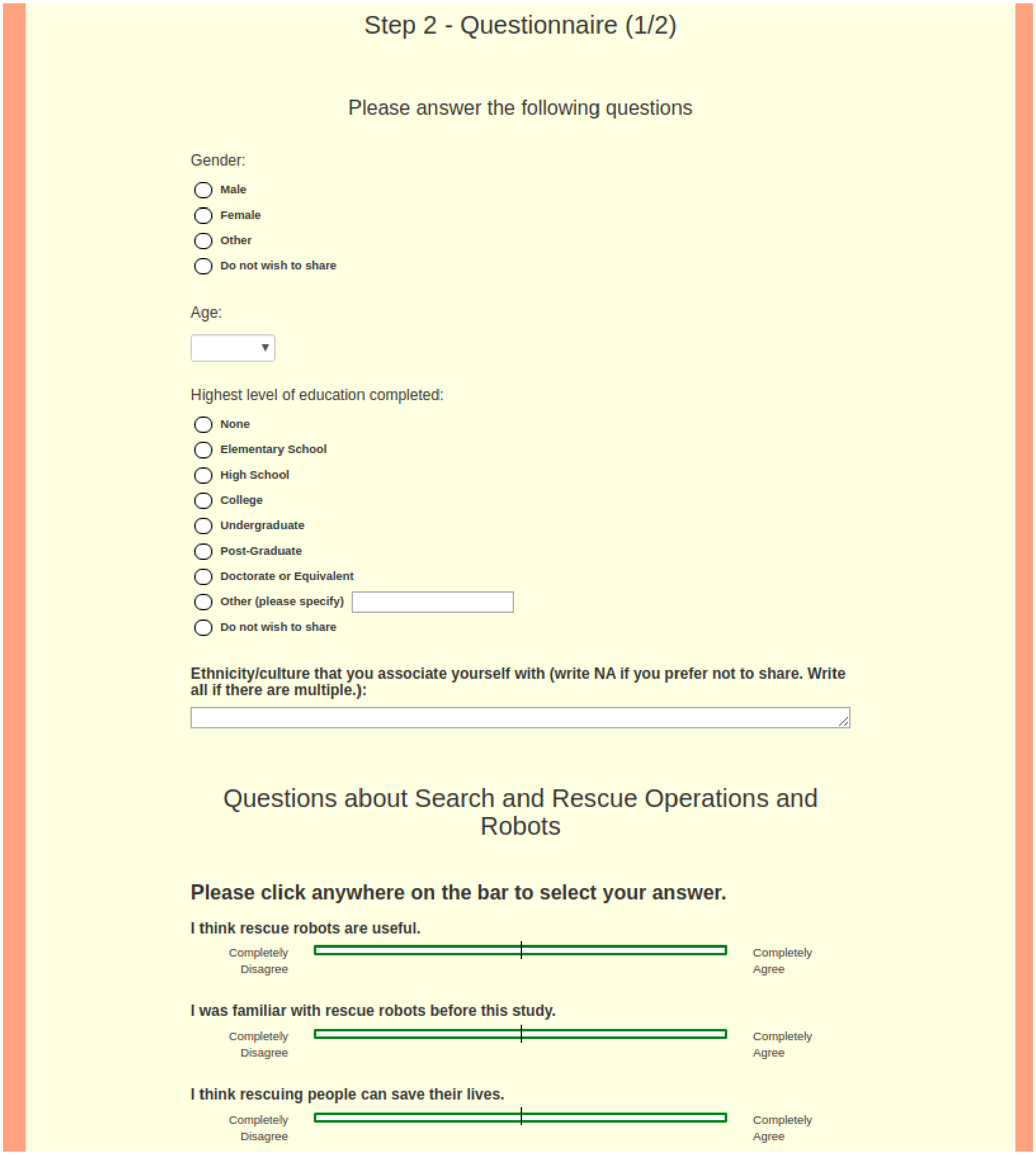


Figure B.4: Main task where participants were asked to map given sentences related to SAR with emotion(s)



Step 2 - Questionnaire (1/2)

Please answer the following questions

Gender:

- Male
- Female
- Other
- Do not wish to share

Age:

Highest level of education completed:

- None
- Elementary School
- High School
- College
- Undergraduate
- Post-Graduate
- Doctorate or Equivalent
- Other (please specify)
- Do not wish to share

Ethnicity/culture that you associate yourself with (write NA if you prefer not to share. Write all if there are multiple.):

Questions about Search and Rescue Operations and Robots

Please click anywhere on the bar to select your answer.

I think rescue robots are useful.

Completely Disagree — Completely Agree

I was familiar with rescue robots before this study.

Completely Disagree — Completely Agree

I think rescuing people can save their lives.

Completely Disagree — Completely Agree

Figure B.5: Survey questions answered at the end of the study

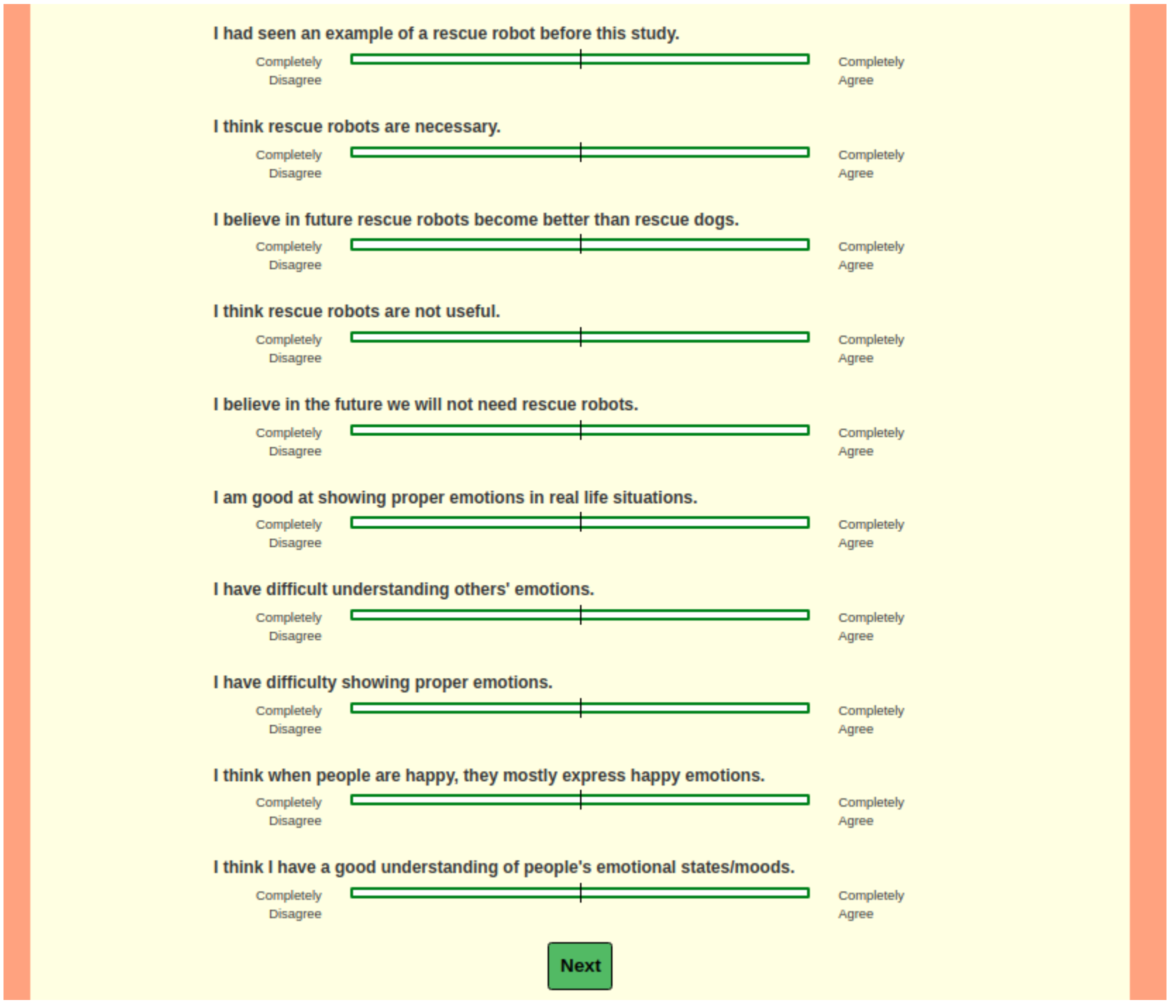


Figure B.6: Survey questions answered at the end of the study

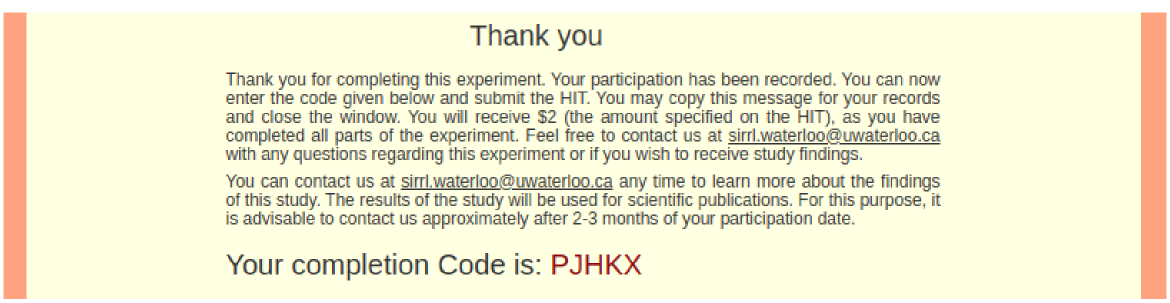


Figure B.7: Completion page thanking participants for completing the study

Appendix C

Online Interface in Experiment 2

Online interface employed in Experiment 2 is given in this appendix.

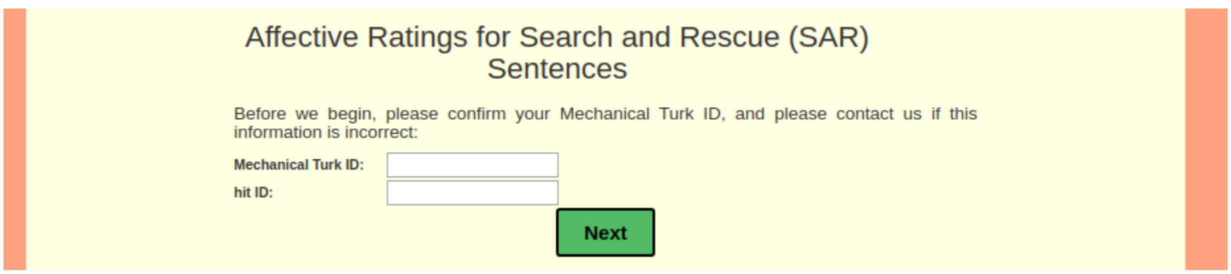


Figure C.1: Login Page

Information and Consent Form

Date: July 2020

Study Title: Affective Ratings for Search and Rescue (SAR) Sentences

We are all from the University of Waterloo in Canada.

Principal Investigator: Prof. Kerstin Dautenhahn, Department of Electrical and Computer Engineering/Systems Design Engineering

Student Investigator: Sami Alperen Akgun, Department of Systems Design Engineering

Collaborator: Dr. Moojan Ghafurian, School of Computer Science

Overview: In this study, we are going to show you some sentences/words related to situations happening in Search and Rescue (SAR) area, and we want to know **how good**, **how active**, and **how powerful** you think these sentences are.

For example, "mother" is generally perceived to be very good, quite powerful, and quite active. On the other hand, "an infant" is perceived to be very good, but quite powerless and only slightly active.

These kind of ratings can also be applied to behaviours. For example, "the action of helping others" is considered to be very good, very powerful and quite active, while "lying to people" is perceived to be very bad, slightly powerless and somehow inactive.

Procedure: The procedure of the task is as follows: (1) you will first sign the consent form and (2) read the instructions. (3) you will rate different sentences/words on three aspects regarding how good, how active and how powerful you think these sentences.

If you do not wish to rate certain sentences, you may choose to not answer them. You can do so by contacting us at sirrl.waterloo@gmail.com. **You can complete this HIT only once**.

Duration/Time: The whole process will take around 5 minutes.

Confidentiality: You will not be identified individually in any written reports of this research. However, your answers in open text form might be used as anonymous quotations. Note that when information is transmitted over the internet, privacy cannot be entirely guaranteed. There is always a risk that your responses may be intercepted by a third party (e.g., government agencies, hackers). Any confidential information (namely your MTurk ID) will be removed from our records within 1 year. You may withdraw your consent and your data within this period by contacting us. The data will be securely stored in locked offices in the research laboratory of Prof. Kerstin Dautenhahn (to which only she has access. Your identity will be kept confidential), and on the principal investigators' computer for analysis. Anonymized data may be published publicly in the future. Anonymized data will be stored for a minimum of 7 years.

Right to Ask Questions: Please contact us at sirrl.waterloo@gmail.com with questions or concerns about this study. This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE# 41900). If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

Payment for participation: You will be paid $0.3 if you complete all the steps of the experiment. If you do not wish to continue at any point, your participation in the study will end and you will be still paid $0.3. If you wish to stop at any time, please **do not submit the HIT** and contact us at sirrl.waterloo@gmail.com.

To accept your HIT, you must add the code that is provided to you on the end page, or given to you by us if you wish to end the study earlier. **We will not be able to accept your HIT if you do not provide the correct code.**

Voluntary Participation: Your decision to take part in this research is voluntary. You do not have to answer any questions you do not want to answer. You can stop at any time. If this is the case, please exit the page and abandon the Mechanical Turk HIT. We ask you to contact us in this situation. In this case your data will be discarded and not be used in the study.

Thank you for your interest in our research and for your assistance with this project. You must be 18 years of age or older to consent to take part in this research study. By accepting this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities. With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.

I Agree

Figure C.2: Information and Consent Form

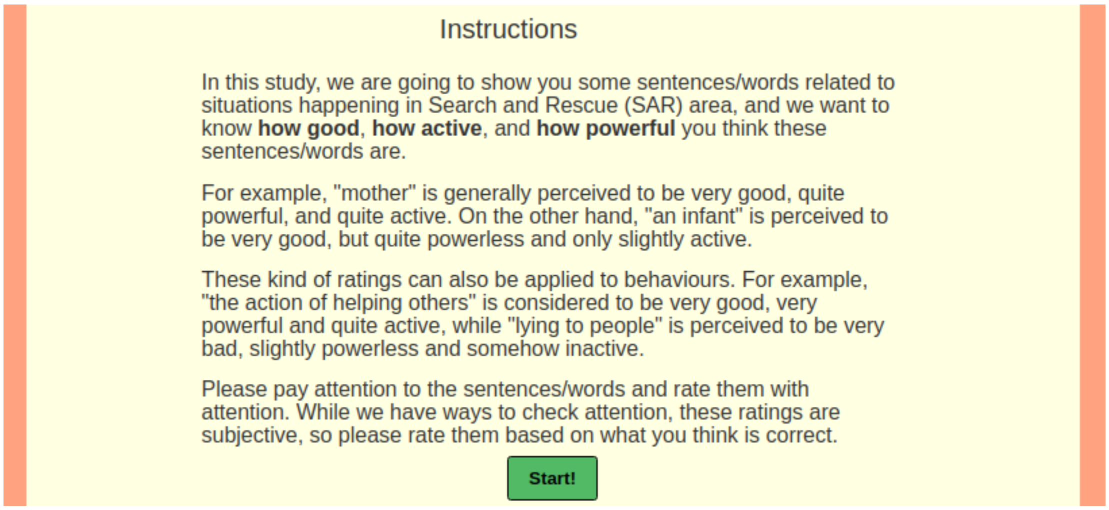


Figure C.3: Instructions

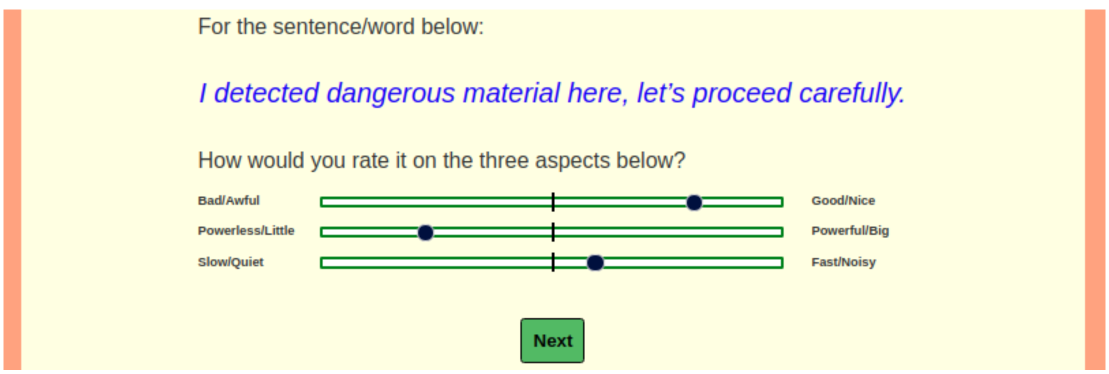


Figure C.4: Main task asking participants to rate given sentences/words in three different scales (EPA dimensions)

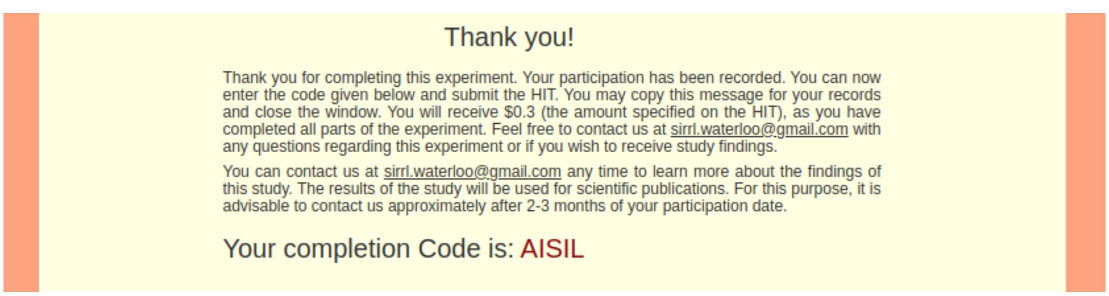


Figure C.5: Completion page thanking participants for completing the study

Appendix D

Online Interface in Experiment 3

Online interface employed in Experiment 3 is given in this appendix.

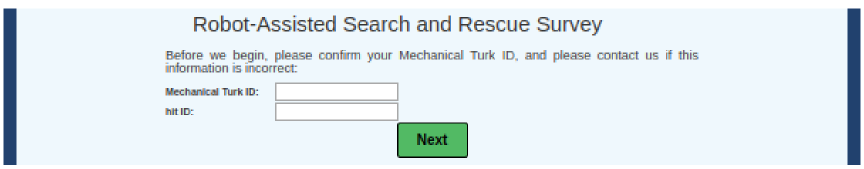


Figure D.1: Login Page

Information and Consent Form

PLEASE USE GOOGLE CHROME FOR THIS STUDY!

Date: April 2021

Study Title: Evaluation of Common Situations Occurring During Search and Rescue Missions

We are all from the University of Waterloo in Canada.

Principal Investigator: Prof. Kerstin Dautenhahn, Department of Electrical and Computer Engineering

Co-Principal Investigators:

- Prof. Moojan Ghafurian, School of Computer Science
- Prof. Mark Crowley, Department of Electrical and Computer Engineering

Student Investigators:

- Sami Alperen Akgun, Department of Systems Design Engineering
- Hamza Mahdi, Department of Electrical and Computer Engineering
- Shahed Saleh, Department of Mechanical and Mechatronics Engineering

Overview: You are invited to participate in a study. We are doing research on usage of robot called Husky in search and rescue missions. Before you agree to take part in this study, it is important for you to understand what you will be asked to do. Please read the information below and make a decision about whether or not you wish to participate. Feel free to ask us any questions about the study, or any part of this form that may not be clear.

In this study, we aim to ask your opinion about the common situations occurring during search and rescue missions. We will first show you some emotions implemented on a robot. Then, we will show you short videos of a robot, and ask you to assume that the robot is helping you in a search and rescue mission and want to let you know about a situation. Then, we will ask you questions about it. Lastly, you will answer a few sets of questions about you and your experiences. Your answer to the questions will be logged. This does not require you to download anything onto your computer, however, you need to be able to load a few short videos to watch them.

The outcome of this study will enable us to create more efficient search and rescue robots.

Procedure: The study will consist of three parts: (1) getting familiar with emotions of a robot, (2) watching videos and answering questions about them, and (3) responding to a questionnaire. We prefer to receive full responses as they provide valuable information that can help in our study. However, if you do not wish to answer certain questions, you may choose to not answer them. You can do so by selecting "not sure" or "do not wish to answer" if available, or by contacting us at sirrl.waterloo@gmail.com. **You can complete this HIT only once.**

Duration/Time: The whole process will take up to 30 minutes.

Confidentiality: You will not be identified individually in any written reports of this research. Note that when information is transmitted over the internet, privacy cannot be entirely guaranteed. There is always a risk that your responses may be intercepted by a third party (e.g., government agencies, hackers). Any confidential information (namely your MTurk ID) will be removed from our records within 1 year. You may withdraw your consent and your data within this period by contacting us. The data will be securely stored in locked offices in the research laboratory of Prof. Kerstin Dautenhahn (to which only she has access. Your identity will be kept confidential), and on the principal investigators' computer for analysis. Anonymized data may be published publicly in the future. Anonymized data will be stored for a minimum of 7 years.

Right to Ask Questions: Please contact Sami Alperen Akgun or Moojan Ghafurian at sirrl.waterloo@gmail.com with questions or concerns about this study.This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE# 43033). If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

Payment for participation: You will be paid \$3 if you complete all the steps of the experiment. If you do not wish to continue at any point, your participation in the study will end and you will be remunerated for the proportion of what you completed (number of questions that you have answered). If you wish to stop at any time, please **do not submit the HIT** and contact us at sirrl.waterloo@gmail.com. In that case, you will received the amount specified on the HIT plus an additional payment based on the number of questions that you have completed. For example, if you wish to stop in the middle of the study, you will receive (\$1.5)+(\$1.5/2)= \$2.25.

To accept your HIT, you must add the code that is provided to you on the end page, or given to you by us if you wish to end the study earlier. **We will not be able to accept your HIT if you do not provide the correct code.**

Voluntary Participation: Your decision to take part in this research is voluntary. You do not have to answer any questions you do not want to answer. You can stop at any time. If this is the case, please exit the page and abandon the Mechanical Turk HIT. We ask you to contact us in this situation, so we can remunerate you for the proportion of what you completed. In this case your data will be discarded and not be used in the study.

Thank you for your interest in our research and for your assistance with this project. You must be 18 years of age or older to consent to take part in this research study. By accepting this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities. With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.

I Agree

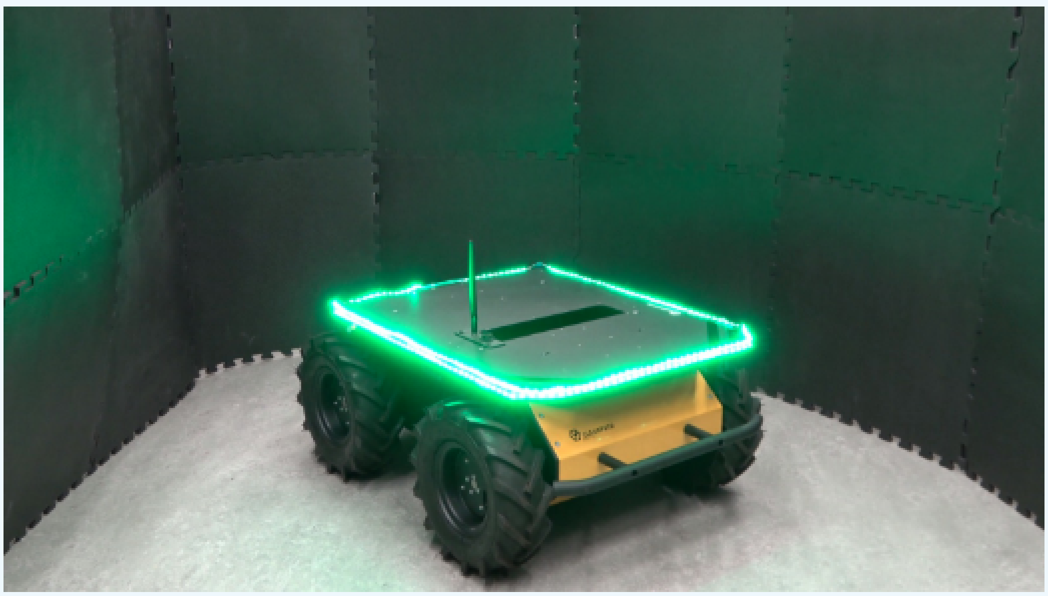
Figure D.2: Information and Consent Form

Instructions

PLEASE USE GOOGLE CHROME FOR THIS STUDY!

Step 1 - Getting Familiar with the Husky robot and its Emotional Expressions

In this step, we will show you some emotional expressions implemented on the Husky robot using LED lights and we will ask you to remember these patterns. Here is **an example screenshot** from the video that you will watch in this step:



After watching the videos of Husky's emotional expressions and responding to questions about them, you will be moving to the next step of this study, which will be explained later.

Let's start learning emotion patterns of the Husky robot by clicking **Next** button.

Next

Figure D.3: Instructions for emotion training

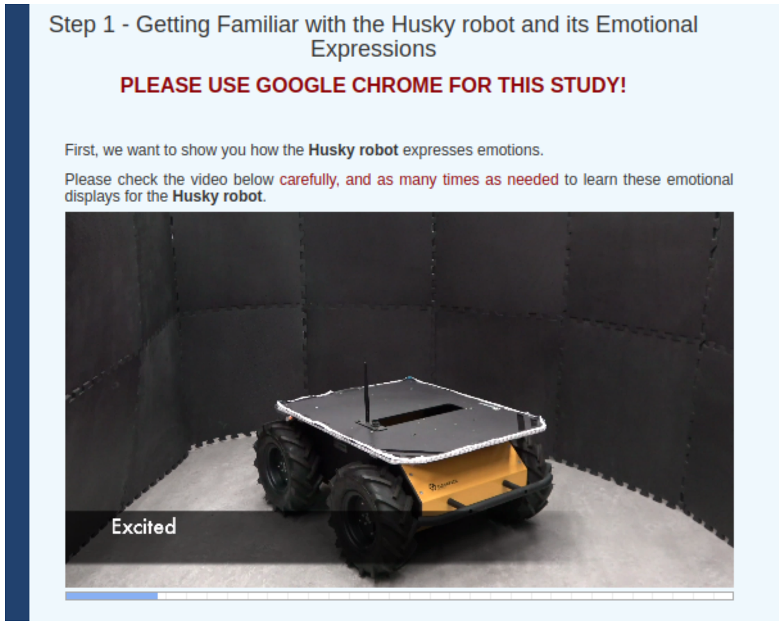


Figure D.4: Emotion training video showing all implemented emotions

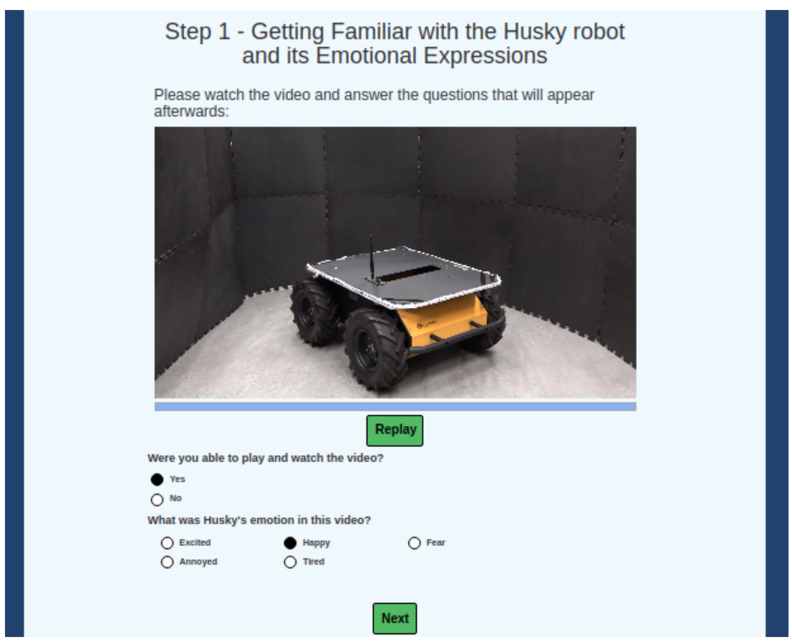


Figure D.5: Emotion Training Test Step which participants see all emotions implemented step by step, and they were asked to choose correct emotion

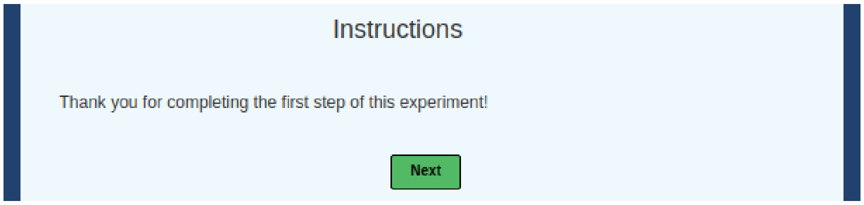


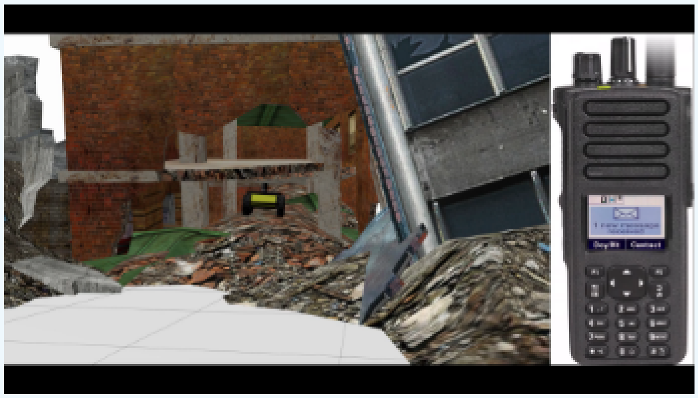
Figure D.6: Completing Emotion Training Part

Figure D.7: Instructions for the main task and introduction video showing disaster area as well as Husky itself (Page 1)

Instructions (2/2)

Step 3 - Evaluating the Videos:

The **Husky** robot will notify you about what is happening in the area. For each message that the robot wants to convey to you, you will watch two videos. The first video shows **Husky moving in the environment and notifying you about a specific situation relevant to the search and rescue mission.** You will be notified that you have received a text message from Husky via your radio device. **At this step, you cannot see the messages.** The video will look like this:



The second video will show Husky alongside the text message that it sent to you. Unfortunately, due to low radio coverage in the area, **the text message you received is noisy.** In this video, **on the left, you will see Husky, and on the right the text message sent to your device.**

After watching both videos, you will be given a list of possible messages:

- I can again communicate with our team outside of the building
- I lost communication with our team outside of the building
- I am stuck and might need help to proceed
- I detected dangerous material here, let's proceed carefully
- I believe we are behind schedule. I also noticed it is getting dark and there is not much time left
- I found an item that could belong to a person. Maybe the person is nearby
- My battery is running low and I will stop working soon
- I think I found a surviving person
- I detected that there might be a risk of further collapse so we should only proceed with caution
- I think I heard someone is calling for help, we might have found a survivor

Your task will be to choose the message what message Husky wanted to convey to you considering **the noisy text message, emotional expressions of Husky and the scenario itself (i.e., the robot's movements in the environment).**

Step 4 - Questionnaire:

After watching videos related to 10 messages and choosing your responses, you will be asked to complete a Questionnaire with a set of questions about yourself.

You should not press "back" on your browser at any point. The study will stop if you press back. Contact us if you want to go back and change your answers.

You need to pay attention to all parts of this study and provide honest answers.

Figure D.8: Experiment 4 - Instructions for the Main Task (Page 2)

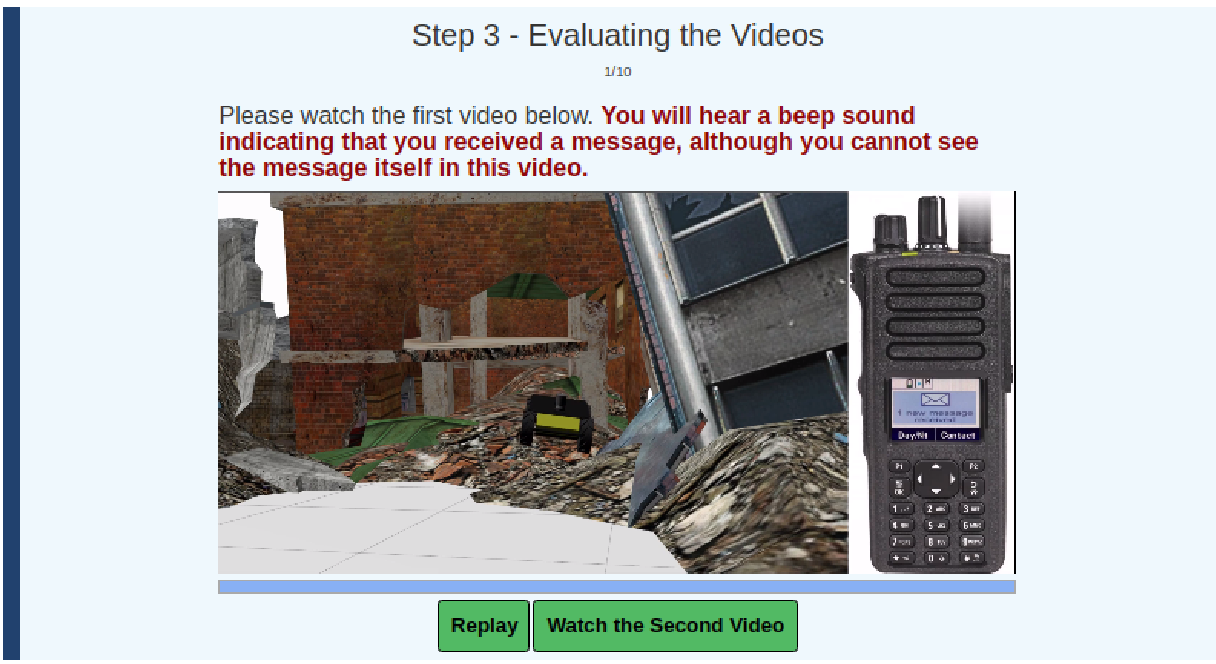


Figure D.9: Main Task (Page 1)

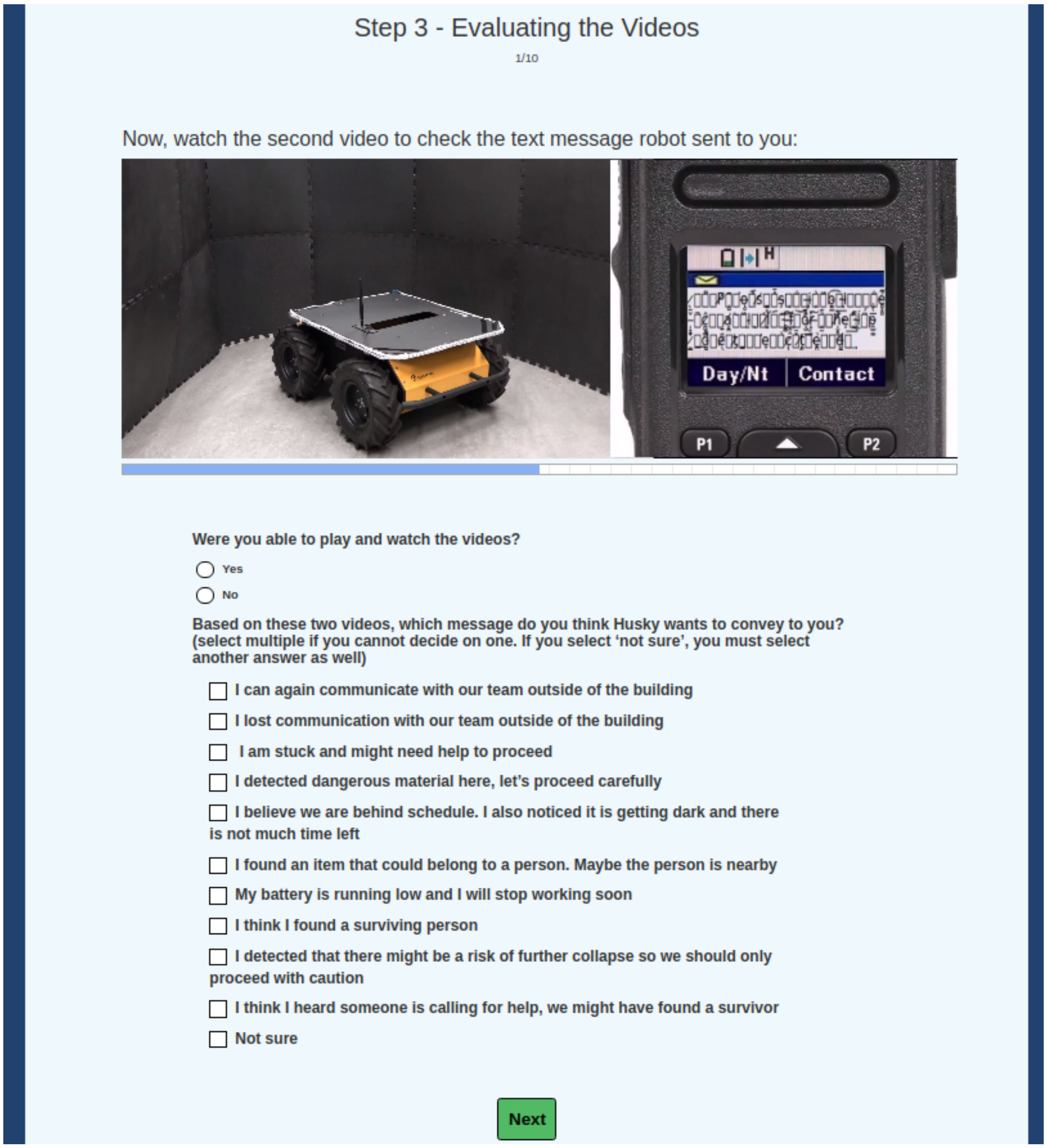
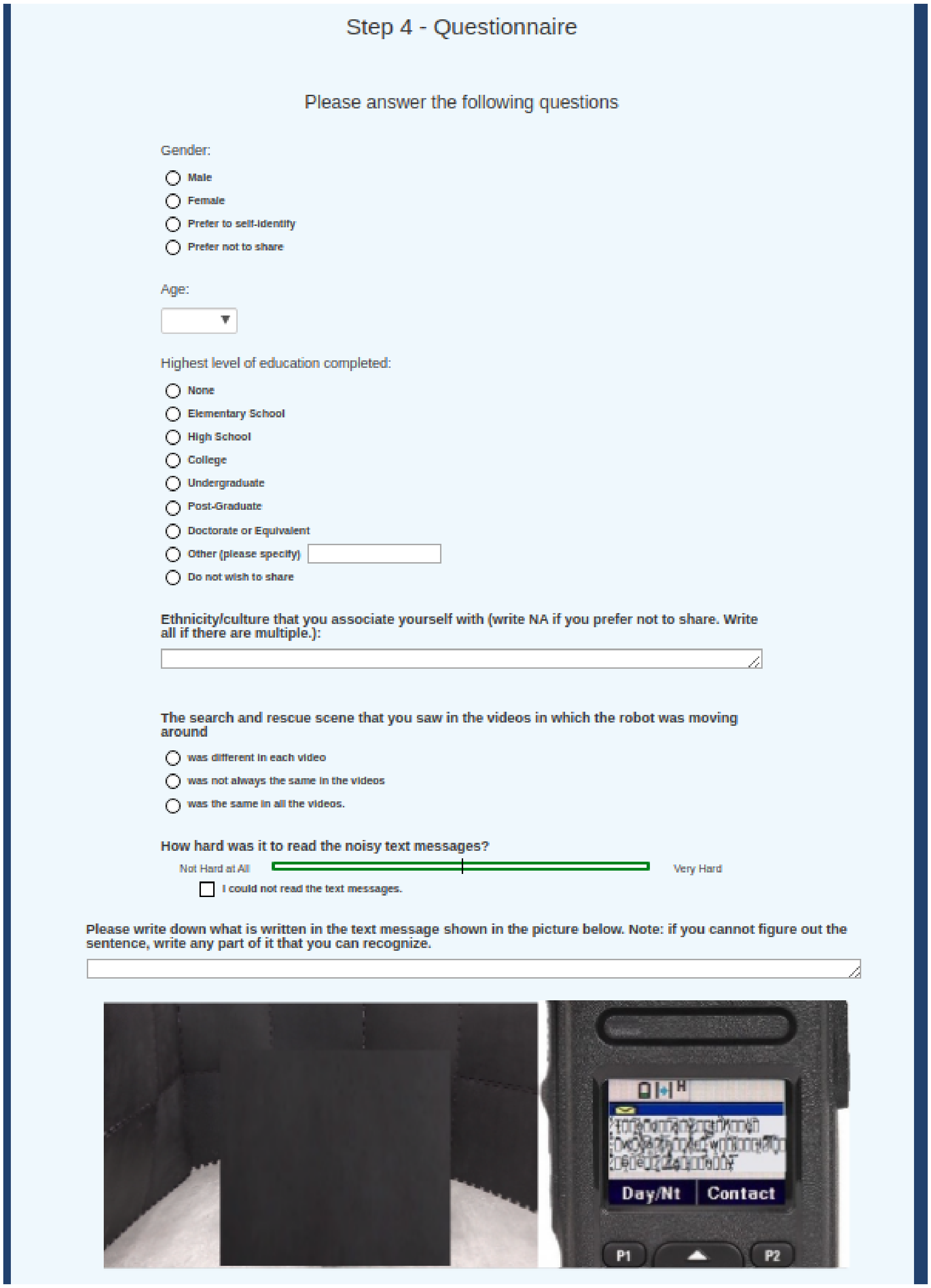


Figure D.10: Main Task (Page 2) that participants asked to select the message Husky wants to convey based on two videos they saw



Step 4 - Questionnaire

Please answer the following questions

Gender:

- Male
- Female
- Prefer to self-identify
- Prefer not to share

Age:

Highest level of education completed:

- None
- Elementary School
- High School
- College
- Undergraduate
- Post-Graduate
- Doctorate or Equivalent
- Other (please specify)
- Do not wish to share

Ethnicity/culture that you associate yourself with (write NA if you prefer not to share. Write all if there are multiple.):

The search and rescue scene that you saw in the videos in which the robot was moving around

- was different in each video
- was not always the same in the videos
- was the same in all the videos.

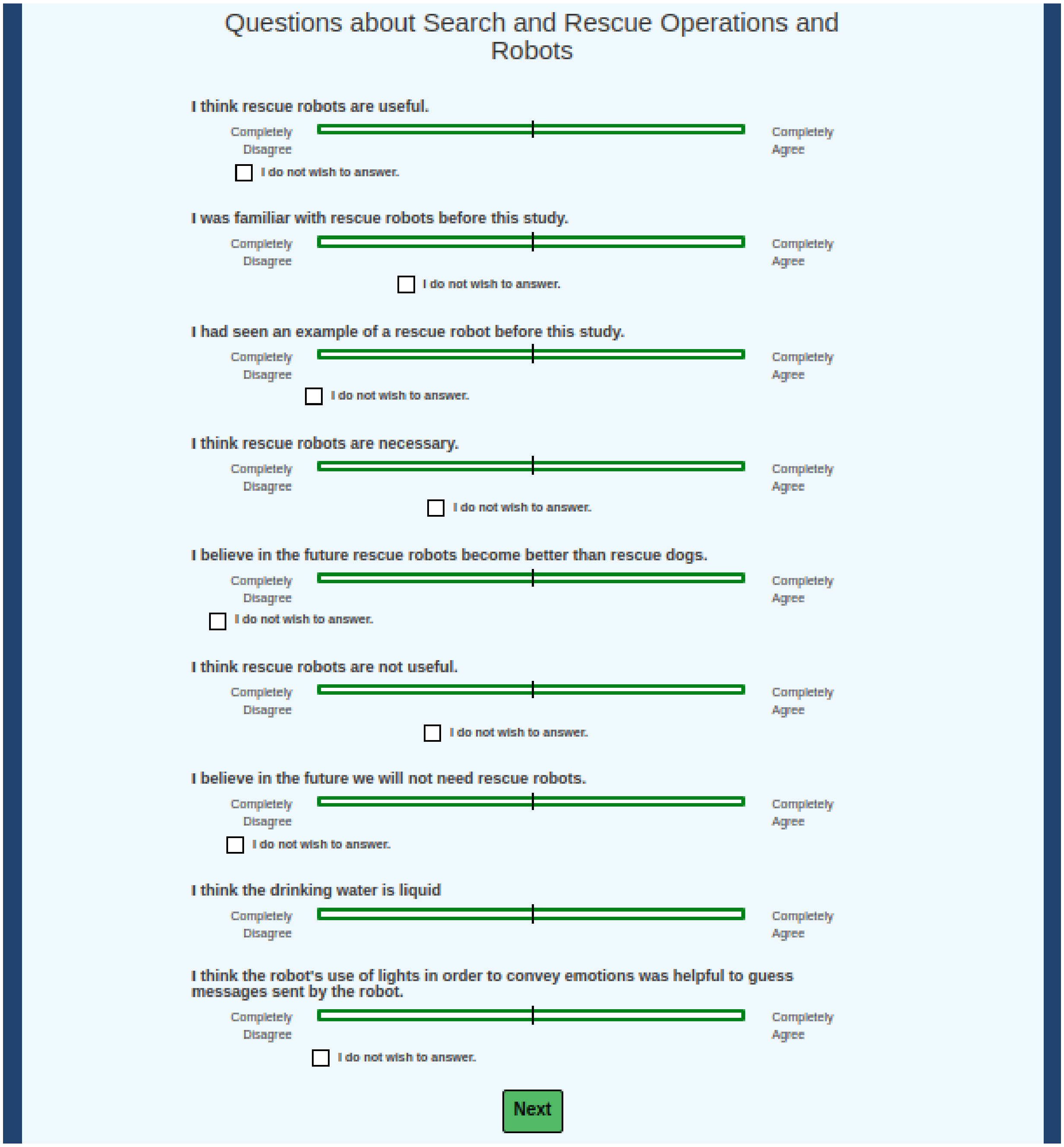
How hard was it to read the noisy text messages?

Not Hard at All — Very Hard

- I could not read the text messages.

Please write down what is written in the text message shown in the picture below. Note: if you cannot figure out the sentence, write any part of it that you can recognize.

Figure D.11: Survey questions answered at the end of the study



Questions about Search and Rescue Operations and Robots

I think rescue robots are useful.

Completely Disagree — Completely Agree

I do not wish to answer.

I was familiar with rescue robots before this study.

Completely Disagree — Completely Agree

I do not wish to answer.

I had seen an example of a rescue robot before this study.

Completely Disagree — Completely Agree

I do not wish to answer.

I think rescue robots are necessary.

Completely Disagree — Completely Agree

I do not wish to answer.

I believe in the future rescue robots become better than rescue dogs.

Completely Disagree — Completely Agree

I do not wish to answer.

I think rescue robots are not useful.

Completely Disagree — Completely Agree

I do not wish to answer.

I believe in the future we will not need rescue robots.

Completely Disagree — Completely Agree

I do not wish to answer.

I think the drinking water is liquid

Completely Disagree — Completely Agree

I think the robot's use of lights in order to convey emotions was helpful to guess messages sent by the robot.

Completely Disagree — Completely Agree

I do not wish to answer.

Next

Figure D.12: Survey questions answered at the end of the study

Thank you!

Thank you for completing this experiment. Your participation has been recorded. You can now enter the code **WDOIH** and submit the HIT. You may copy this message for your records and close the window. You will receive $1.5 bonus (in addition to the amount specified on the HIT), as you have completed all parts of the experiment. Feel free to contact us at sirrl.waterloo@gmail.com with any questions regarding this experiment or if you wish to receive study findings.

You can contact us at sirrl.waterloo@gmail.com any time to learn more about the findings of this study. The results of the study will be used for scientific publications. For this purpose, it is advisable to contact us approximately after 2-3 months of your participation date.

Figure D.13: Completion page thanking participants for completing the study