UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

Musical utterances to evoke empathy and prosocial behavior toward a hospital robot

Theresa Höfker

M.Sc. Thesis
June 2023

Supervisors:
dr. K. P. Truong
dr. B. R. Schadenberg
dr. ir. E. C. Dertien

University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Abstract

Semantic-Free Utterances can be an effective way of communication in human-robot interaction to convey intentions and emotions to the users in simple and easy-to-understand ways. Although a lot of research is done on semantic-free utterances, musical utterances are a type of semantic-free utterances that are barely researched beyond expressing emotions. In the context of a transportation robot in a hospital needing help, they seem to be promising to evoke empathy and therefore elicit prosocial behavior from people in the surrounding area of the robot. However, to date, it is not known yet whether they are successful in doing so. Therefore, our study looks at the extent musical utterances are able to communicate intentions and emotions that result in empathy and prosocial behavior toward a robot. Together with sound designers, we created musical utterances for specific scenarios in which the robot has a problem where it needs support from people. We tested these sounds in an online video study where participants watched videos of the robot in situations where it is stuck and either used no sounds, beeping sounds, or musical sounds. Our results indicate that people hearing the musical utterances as means of communication by the robot perceived more empathy evoking emotions. Interestingly, the musical utterances evoked higher levels of cognitive empathy compared to no sounds and beeping sounds. Against our expectations, the confidence of the participants in the interpretation of the robot's intentions was the highest in the beeping sound condition. For all other cases, our results did not indicate significant differences between the sound conditions. Therefore, it can be concluded that for the communication of emotions musical utterances seem useful, but to make the robot's intention more legible, it is more effective to rely on simpler sounds like beeping sounds. Still, especially with regard to empathy and prosocial behavior, we cannot draw concrete conclusions and future research should consider this.

Acknowledgments

To realize my research and be able to present the work in this thesis, many people were involved in the process. Without them, my work would not have been possible.

I would like to thank my supervisors Khiet Truong and Bob Schadenberg for their continuous support. They motivated me to create a project I really enjoyed working on. During our weekly meetings, they were always enthusiastic about my work and took the time to share their knowledge and insights. They always had an open ear for me throughout out all of the ups and downs of a thesis. I truly enjoyed working with them.

Importantly, without the sound designers, I would not have been able to carry out this study. They devoted a lot of their time to creating the musical sounds and were at all times very enthusiastic about it. Also, afterward, they were always available to answer all my questions regarding the sound design. I am very thankful for the efforts they took and the contributions they made to this research.

Similarly, Edwin Dertien and Benno Spieker were of great help in the initial phase of the thesis. Their knowledge and insights into the field of music and robotics helped me to figure out the path for my thesis. Also, with the help of Edwin, I was able to make use of a robot which would have been a lot harder otherwise.

Finally, I am grateful for the support I received from my friends, housemates, and family. Whether it was by acting in the videos, giving feedback, or showing moral and mental support, I could not have done it without them.

Contents

Chapter 1

Introduction

In hospitals, a large variety of transportation tasks are executed. Although these processes are often invisible to patients, they can have a great impact on their experience. One example is the flow of materials. Within a hospital, a great amount of materials needs to be organized. Therefore, a complex transportation system to organize the flow of materials exists in hospitals [1]. This and other tasks are currently executed by humans. However, hospital personnel taking over these tasks could pose great challenges for the future like staff shortage and exposure to risks. By 2030, a shortage of 15 million healthcare workers worldwide is expected [2]. Next to that, transporting materials carries additional risks for the workers. For example, during the COVID-19 pandemic, healthcare workers were exposed to the virus and infections with it led to further staff shortage [3]. Taking this together, the healthcare system is facing and will face more and more challenges that need to be solved to ensure a functioning system.

Burdens for the healthcare system, like staff shortages and exposure to risks, can be counteracted by expanding the use of technologies. By relying more on technologies in this sector, the efficiency of healthcare workers can be increased. They can focus on the tasks which essentially need humans to execute them [2]. In addition, the use of technology gives healthcare workers the chance to focus more on the tasks they enjoy doing. One example of introducing technology in the healthcare sector is the use of robots. Accordingly, the Harmony project[1] aims to develop a robot that executes on-demand delivery tasks and automates the bioassay sample flow, so the flow of test samples. These are tasks that are currently executed by staff members who are overqualified for these [4].

The goal of the project is to develop a robot that can navigate autonomously through the hospital. However, still, situations might occur in which the robot requires some form of support or help from people in its environment. An example is when a door that the robot needs to go through is closed. However, it is very likely that

[1] The website of the project can be found here.

experts on the robots are not nearby and might not even be needed to resolve those problems. Hence, people in the environment of the robot should engage in prosocial behaviors to quickly resolve the issue. Resulting, the processes of the robot are not much interrupted and the robot is actually of added value to the hospital staff. Therefore, the robot needs to interact with people in such a way that they are, on the one hand, willing to help and cooperate with the robot and, on the other hand, also understand what the robot needs. When interacting with people, social capabilities are of great importance as robots become part of our social environment. Based on these, interactions between humans and technology should feel as natural as possible to the users. This means, that the machines should become social agents, someone or something that possesses social abilities and capabilities [5]. Although, the project might not aim to develop and design a social robot in the first place, as soon as robots start interacting with people social skills become important.

One example of these social capabilities is experiencing and showing empathy. In short, empathy can be defined as the ability to perceive and understand another's emotions [5]. Empathy is an essential factor in human-human interactions relating to many other variables. For example, a strong relationship exists between empathy and forgiveness [6]. People that are more empathetic are also more easily able to forgive others. Similarly, empathy and trust are also closely related [7]. People that show more empathy are trusted more by others. These are only two examples of why empathy is an important interpersonal factor.

Most interesting for the current use case is the relationship between empathy and prosocial behavior. The more empathetic someone is the more likely they are to engage in prosocial behavior. The results from Welp and Brown [8] for example show that by feeling empathy for someone in need of help, an altruistic motivation to help this person is elicited. This relationship between empathy and prosocial behavior is highly relevant in the current use case where the robot possibly will encounter situations in which it wants prosocial behavior from the people around it, such as a closed door or another obstacle in the way. Hence, the robot should communicate in such a way that it evokes empathy and prosocial behavior.

To evoke empathy, it is important that the target of empathy communicates emotions. Displaying emotions clearly and communicating them, increases the extent to which people feel empathetic [9]. Voice is one important modality to communicate emotions in human-human interaction [10]. Interestingly, non-semantic signals are at least similarly important to display emotions. In speech, the variation of prosodic features closely relates to the display of emotions [11]. Hence, to evoke empathy, emotions should be communicated through the voice. Here, non-semantic information plays a big role to convey these emotions.

In the field of robotics, semantic free utterances (SFU) are a new way of communication that makes use of non-semantic sound signals. In the past years, SFU gained a lot of attention as means of communication [12]. Films like WALL-E and Star Wars with the robot R2D2 demonstrate well that these robots can communicate effectively with and to humans by using sounds without semantic content. Similarly, Zaga et al. [13] demonstrated in their study, that gibberish speech, one form of SFU, is well understood by children and that they were able to communicate the intentions of a robot. The use of sounds had a great additional value compared to no sounds [13]. Nonetheless, the interpretation of SFU is not always straightforward. The interpretation of the sounds is often influenced by other modalities, for example, body movements or facial expressions. Accordingly, people interpret robots as a whole [14]. Next to that, interpreting the subtleties of SFU is hard for people. For example, with regards to emotions, children categorized emotions always in categories of general emotions like happy or sad. However, they did not make more specific differentiations [14]. Accordingly, SFU are a very promising field for robot communication, but there are also still a lot of challenges that are not solved yet.

Musical utterances are one type of SFU that are not well-researched. Musical utterances are musical pieces that result from the variation of different attributes of music, like rhythm, dynamics, timbre, or pitch [12]. Musical utterances have the advantage of being genderless. By relying on genderless robots, negative influences of gender-based biases and stereotypes are avoided. Importantly, voice is one of the important features to communicate gender and hence to make genderless accordingly [15]. Compared to speech-like utterances like gibberish speech, musical utterances are by default genderless and hence avoid any implicit or explicit influence of gender on the perception and interaction. Additionally, musical utterances have the advantage of not being used by any other real or fictional species or characters. The use of music and especially, electroacoustic music, opens up the possibility of creating new ways of expressing utterances. Hence they are not yet charged with expectations from known characters making use of them [16]. In this regard, musical utterances possibly enable new forms of human-machine interactions.

To communicate emotions, musical utterances can be built upon the knowledge already existing. It is long known that musical pieces can convey a lot of emotions to the listeners. Jee et al. [17] made use of these insights. They designed sounds for a robot to communicate emotions and their results showed that people were well able to identify the emotions. Similarly, in another study, researchers developed an approach to generate emotional musical sequences in response to input from human language. The authors argue that their complete system increases user engagement [18]. In various studies, it has been generally shown that people can feel empathetic towards robots (see for example [19] [20]). Also, the results of the study by Holm et al. [21] confirm that sound signals can evoke empathy and prosocial behavior. However, they also stress the importance of looking at musical signals as these are positioned between beeps and a voice. However, the question arises whether people feel empathetic toward robots communicating emotions through musical utterances. To our knowledge, this is not yet researched but musical utterances have several advantages, especially in the current use case. Additionally, Savery et al. [18] also raise the need to further expand the use of musical utterances beyond purely communicating emotions. Accordingly, it is not clear yet whether musical utterances can communicate intentions beyond emotions.

1.1 Report Outline

Taking everything together, we see a gap in the literature concerning how musical utterances can convey the intentions of robots. Currently, musical utterances have been only regarded as means to communicate emotions without conveying a specific need or attempting to elicit a specific social behavior. However, other types of SFU have been shown to effectively communicate the intentions of the robot as well. In addition to this, the relationships between communicating emotions, evoking empathy, and prosocial behavior found in human-human interaction are not yet researched in human-robot interaction (HRI). Especially in the current use case of the Harmony project, they seem of important added value in situations where the robot is stuck and non-experts should engage in prosocial behaviors to resolve the issue. Therefore, in our study, we will look at the following overall research question:

To what extent are musical utterances able to communicate intentions and emotions that result in empathy and prosocial behavior toward a robot?

We designed musical sounds for the communication of the robot. These sounds were tested in an online video study. We investigated to what extent musical utterances are able to be understood and increase the levels of empathy and prosocial behavior.

We structure the report as follows. In Chapter 2[2], we describe relevant background research regarding empathy, SFU, legibility, and prosocial behavior. Based on this we define our concrete research questions and hypotheses. Following this, Chapter 3 is about the sound design that was done to attain musical utterances that fit the current use case well. In Chapter 4, the description of the study to evaluate the sounds can be found and the results of this are described in Chapter 5. The

[2]Parts of the Introduction and Background Chapters are based on the Research Topics report, which has been written by these authors in preparation for this Thesis.

discussion and interpretation of these results follow in Chapter 6. Finally, in Chapter 7 we draw final conclusions and outline recommendations for future studies.

Chapter 2

Background

2.1 Semantic Free Utterances

As described before, one way of expressing emotions is through auditory signals. However, communication by verbal expressions does not necessarily always take place in the form of (spoken) language. Previous research demonstrated that SFU can be well understood by users. SFU are expressions of emotion or intent through auditory communication that include sounds and vocalizations that do not contain any semantic meaning [12]. Generally, SFU is an area that is becoming more important in HRI.

Research distinguishes between four different types of SFU. The first one is gibberish speech. Gibberish speech describes the "vocalization of meaningless strings of speech sounds" [22, p. 163]. The speech is correct in many ways, for example phonetically, syntactically, or grammatically, but does not make sense or contain any meaning. Secondly, non-linguistic utterances are sounds that are not speech and are implemented as feedback mechanisms and social cues in HRI [23]. The third category of SFU is paralinguistic utterances. These are utterances containing the vocal factors of speech by going beyond the content of the utterance. Examples of paralinguistic utterances are laughter or moans [12]. Lastly, musical utterances are a form of robot communication that results from the variation of different attributes of music like rhythm, dynamics, timbre, or pitch [12]. The sounds are deliberately created for the communication of intent or state of robots [17]. All four categories of SFU are distinct from each other. However, what unites them is their ability to convey information to users without utilizing words.

SFU have many advantages as means of communication compared to other communication channels. First of all, they are non-linguistic and non-cultural. In the hospital, the chosen context for the study, a diverse group of people possibly encounters the robot. By communicating through SFU, robots are no longer dependent on a specific language or dialect. Accordingly, they are considered advantageous in multilingual and multicultural contexts [12]. Additionally, the expectations of the users can be more easily managed. Often, when robots make use of complex, human-like speech, the users have higher expectations of the robot with regard to its capabilities [12]. By using SFU these expectations can be better managed. However, one drawback of SFU is that they are often context-dependent. People interpret SFU in relation to what they perceive through other channels. For example, in one study, the interpretations of the emotions of a robot were based on context rather than non-linguistic utterance. The non-linguistic utterances were interpreted in the light of the bigger context [24]. Similarly, the interpretation of SFU is also influenced by the embodiment of the robot. Depending on the sound-embodiment match, non-linguistic utterances are interpreted more or less emotion-arousing [25]. However, when taking these drawbacks into account, SFU can be a promising field within HRI.

A lot of research has been done concerning how well SFU are understood by users. However, as described earlier, from the world of animation it already became clear that robots can communicate without the use of words. Wall-E and R2D2 make use of beeps, whirrs, and squeaks to engage in interactions. Their communication is highly effective [12]. Another example from research is the paper by Lee et al. [26]. They argue that non-linguistic utterances can be easily understood by the users of robots. The understanding of non-linguistic utterances is more natural and therefore simpler and faster. One reason for this is that non-linguistic are not as cognitively demanding. This is also a result of the short processing time needed to understand non-linguistic utterances [26]. Also, Breazeal [27] designed vocalizations for the Kismet robot. By using gibberish speech, the robot is able to convey emotions and intentions. SFU are able to communicate intentions beyond emotions. However, with regard to musical utterances, barely any research exists about the extent they are able to communicate intentions beyond emotions and therefore seems interesting to further look into.

Taking this together, SFU gain more and more attention in HRI. There are different kinds of SFU that differ in the way they are created and what they sound like. Overall, SFU are non-linguistic and non-cultural and also enable better expectation management with regard to the skills and capabilities of the robot. Hence, in the current use case, they seem to be a good fit as in a hospital the population is very diverse and has different experiences with robots. Previous research has also shown that SFU can convey intentions and emotions to users.

2.1.1 Musical Utterances

The use of musical sounds in HRI might be a promising field within SFU that is not sufficiently researched yet. Musical utterances are deliberately created for the communication of emotions and intentions of robots [17]. This distinguishes musical utterances from music in general as well as musical pieces. That music conveys emotions is nothing new or surprising and it is part of the nature of music. By creating a musical piece, the composers code their own feelings and emotions into musical scores that can be perceived by the listeners. Next to that, here is also the assumption that music and speech emerged from the same evolutionary root from a psychological perspective [28]. Hence, they seem to share many characteristics in conveying information.

Importantly, musical utterances also distinguish themselves from both auditory icons and earcons. Auditory icons are sounds that are used in technologies and which are based on a sound occurring in the real world. They represent virtual objects of the system. Hence, the users are already familiar with the sounds due to their inherent nature [29] and therefore, already attribute meaning to them. One example is the clicking sound when taking a screenshot on the phone resembling the clicking of a camera shutter [30]. Auditory icons are different from musical utterances in this aspect of naturalness. Musical utterances are consciously designed to convey emotion and meaning and do not relate to real-world sounds.

Next to that, earcons are defined as ”nonverbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction” [31, p. 13]. Accordingly, earcons represent objects of technology. Both types of sounds are used as feedback for the activity of the user [30]. This is another difference from musical utterances. With musical utterances, robots also start an interaction and they can, but do not necessarily need, to occur in response to user input. The last difference between auditory icons, earcons, and musical utterances is the duration which is very short for auditory icons and earcons. For example, the majority of auditory icons and earcons designed by Larsson et al. [32] was no longer than one second. However, musical utterances are often a lot longer than this (see for example [17]). Hence musical utterances can be clearly distinguished from the other types of sound communication.

To create and design musical utterances, sound designers can take advantage of well-researched insights into how music can convey emotions. Different features of music, for example, tempo, key, pitch, melody, harmony, and rhythm, relate differently to emotions. Already in 1935, Hevner [33] investigated the relationship between minor and major keys and their affective quality. By his study, he confirmed that musical pieces written in a major key are identified as dynamic, displaying excitement and joy, and sounding happy. In contrast, by making use of minor keys, musical sequences are interpreted as, among others, melancholic, passive, and depressive [33]. This clearly points out that the manipulation of features of music influences how people perceive music. Since then, a lot of research has been done regarding how musical features can be manipulated to display various emotions. For example, Juslin [34] looked into the psychological aspects of emotions in music. Initially, he outlined a number of emotions and how they are represented through musical features (for the full list see [34, p. 335]). This overview is in line with a lot of other research demonstrating that depending on how musical parameters are manipulated and come together, the musical piece is perceived differently by the listeners.

Some research exists on how to implement music in robots to communicate emotions. Jee et al. [35] designed musical sounds expressing specific emotional states of a robot. For example, happiness is expressed through a high tempo, major keys, a higher pitch, consonant harmonies, and a regular rhythm. On the contrary, sadness is signaled through a slow tempo, minor keys, a narrow range of notes, dissonant harmonies, and a firm rhythm [35]. One example of a musical utterance can be seen in Figure 2.1. The aim here is to express sadness. Here, factors like pitch, timbre, and the key of the utterance are manipulated deliberately to convey sadness [17].

Figure 2.1: Sadness expressed through a musical score.

Note. From "Sound design for emotion and intention expression of socially interactive robots" by Jee et al., 2010, *Intelligent Service Robotics*, p. 205.

First studies indicate that musical utterances can be well understood by users of robots. When comparing the impact of musical sounds with the expressivity of facial expressions, Jee et al. [35] found that the participants were more strongly influenced by the musical condition in contrast to the facial expression condition. Hence, music conveys stronger emotions than facial expressions [35]. Similarly, Savery et al. [36] demonstrated that people are well able to recognize emotions from music. In their study, relying only on audio and not on additional gestures elicited the clearest emotion identification [36]. Therefore, when trying to show emotions and elicit emotions in an observer, musical utterances could be very effective in HRI.

Overall, musical utterances can be used to communicate emotions and clearly distinguish themselves from other types of sound communication. When creating musical utterances for HRI, creators can build upon the long-known and well-researched association between music and emotions. To date, a lot of research exists on how the manipulation of musical features relates to concrete emotions. First attempts have been made to implement musical utterances as means of communication in HRI. They have been successful in communicating specific emotions. People were able to understand and interpret the emotions. However, current research has not yet looked at the implications of communicating emotions through musical utterances and whether this can evoke empathy in HRI. Additionally, so far, it is not yet clear whether the communication of intentions beyond emotions can be also achieved through musical utterances. However, trying to answer these questions seems highly relevant to the current use case.

2.2 Legibility

In order for human-robot collaboration and interaction to be successful, the behaviors of the robot have to be legible. In the current use case, we are interested in whether musical utterances make the robot's behavior more legible. In its original meaning, legibility was a property of physically printed materials and referred to the readability of these materials [37]. With regards to robots, legibility concerns whether their behavior is readable. Lichtenthäler and Kirsch [38] defined legible robot behavior as behavior that makes the intentions of the robot clear and understandable and that meets the expectations of the user. However, there seems to be disagreement about whether legible behavior actually aligns with these expectations. Dragan et al. [39] defined this as predictable behavior. Importantly predictability and legibility can also contradict each other. When behavior is legible, so understood by the user, it does not mean it is predictable, so it meets the expectations of the use [39]. However, for the current use case, we ascribe to the first part of the definition of legibility, the understanding of the behavior and the intentions of the robot. Meeting the expectations of the user is not of main interest currently. Hence, the discussion about how predictability and legibility are related is out of the scope of the current research. Accordingly, in the current research, legible robot behavior is regarded as behavior that is correctly understood by the users and which makes the intentions of the robot clear.

Currently, a lot of research about legibility is looking at robot motion and navigation. For example, Angelopoulos et al. [40] considered situations in which people and robots need to cross each other's ways in narrow spaces. Their study showed that using non-verbal behaviors, more specifically deictic gestures improve the legibility of the robot's behavior and intentions [40]. More generally, a diverse set of gestures can improve the legibility of the behavior and intentions of a robot. Any additional gesture that relates closely to the goal of the robot improves the legibility. Complementary gestures are able to clarify the intentions [41]. Hence, adding additional channels of non-verbal communication seems to improve the legibility of robot behavior.

So far, legibility has been mainly manipulated with motion and gestures. Gestures generally improve the legibility of robot behavior and intentions. As described before, another way to communicate the intentions of robots is SFU, more specifically musical utterances. However, to our current knowledge, no research has yet studied at whether sounds can make robot behavior more legible. Therefore, the question arises whether musical utterances are able to improve the legibility of robot behavior and intentions.

2.3 Empathy

When the robot shows emotions and the intentions are legible, users might experience empathy. Empathy is an important concept for human-human interaction and relates to many other important interpersonal concepts. Therefore, as robots enter the worlds of humans empathy also becomes important in HRI. Especially in the current context, where the robot wants people to engage in prosocial behaviors, empathy can enhance the likelihood of this happening.

2.3.1 Definition

The term empathy originated from the German notion of Einfühlung. This concerned the projection of one's self into the object that is perceived [42]. Since then, a lot of research has been done regarding empathy and many definitions and theories have been built around it (see [43], p. 146- 147 for an overview). Based on these, empathy can be described as an emotional response that is initially automatically elicited but can also be controlled in a top-down process. Accordingly, the emotion that is perceived by the person is similar to the experience of the object of attention. However, the emotions are not necessarily exactly the same. Therefore, some kind of self-other distinction is still existing [43]. Taking this together, it clearly points out that there are several aspects underlying empathy. In the following, these are described in more detail. Importantly, the target is the person who shows emotions and evokes empathy, and the observer is the person experiencing and expressing empathy.

First, there is a distinction between cognitive and affective empathy [43]. Affective empathy can be defined as the emotional response to the feelings of the object of perception. According to the authors of [44], affective empathy describes the "ability to experience and share the emotions of others" [44, p. 388]. The person experiences the same or similar emotions as the target. On the contrary, cognitive empathy is the cognitive process taking place when understanding the emotions the target is experiencing [44]. These two concepts are distinct aspects of empathy, which is also supported by the finding of activation of different brain areas depending on which kind of empathy is experienced. However, the two concepts cannot be regarded as completely distinct and interplay with each other. Cognitive empathy is the process by which emotions are formed which are then felt by the person through affective empathy. Hence, the automatically elicited affective empathy can be influenced through conscious cognitive empathy [43]. Accordingly, even when separating the two forms of empathy, the interactiveness should not be forgotten.

Additionally, in order for a person to have an empathetic response, curiosity, and appraisal play a big role, especially in HRI, as will be described in Section 2.3.3. Important for an appropriate empathetic response is the curiosity to understand the point of view of the other in that specific situation [45] [46]. It is important to understand the relationship between the target and the situation they are in. When someone is curious about the situation of the other person, they are motivated to understand the situation correctly. The observer is better able to label the target's emotions correctly and accurately. This again also highlights the interaction between the affective and cognitive components of empathy [45]. Related to these automatic and deliberate processes is the appraisal theory of empathy. This describes that if the observer appraises the situation of the target similarly as the target does, the observer will feel empathy towards the target [47]. Accordingly, if the observer fails to appraise the situation similar to the target, there is a high chance they are not experiencing empathy.

One might get the impression that experiencing empathy is a trait that is stable within a person. To a certain extent, this impression is true. According to the trait view of empathy, the capacity to feel empathy varies from person to person, however, one person always feels a similar amount of empathy. These variances can be explained, by, among others anatomical characteristics, genetic factors, and developmental circumstances. Additionally, certain psychological conditions, like psychopathic or autistic individuals, also influence to what extent people feel empathetic toward others [43]. Nonetheless, empathy can also vary within a person. A great variety of factors concerning the situation someone is in also influence the extent to which someone feels empathetic. These can be, among others, concerning the relationship between the observer or the extent to which a need is perceived [43]. This is called situational empathy [48]. Based on this, although empathy is a stable trait that someone possesses to a certain degree, it also varies depending on the situation. Therefore, the level of empathy felt in a specific moment is a result of the interaction between trait and situational empathy. Although we acknowledge empathy being a trait, the focus of the current research is on situational empathy and how this can be evoked given the nature of the research.

Concluding, empathy is a cognitive as well as affective process where both are important for the experience of empathy. By being curious to grasp the situation of the target it becomes easier to classify the emotions of the target. By appraising the situation similarly to the target, the observer experiences empathy. Generally, empathy is a trait that people possess to a certain extent. Nonetheless, their levels of empathy experienced in specific situations can also vary. This aspect of empathy is of specific interest in the current context. We want to try to evoke more situational empathy through musical utterances.

2.3.2 Evoking Empathy

As the levels of empathy someone experiences can depend on situational factors, we can design situations and HRI with the goal of evoking higher levels of empathy. This again might have further positive influences on the interaction. Generally, it is important that the observer perceives the emotions of the target. When emotions are more clearly displayed, it increases the extent to which an observer experiences empathy toward the target. Hence, the more intense the emotions are experienced and the more salient these emotions are displayed by the target, the greater the empathy felt by the observer [9]. Therefore, when we want to evoke emotions within the observer, it is important that the target communicates their emotions strongly so that these can be easily perceived by the observer.

Importantly, it does not only matter to what extent emotions are shown by the target but the type of emotions also plays a role. Negatively valenced emotions evoke higher levels of situational empathy compared to positively valenced ones or neutral ones. A possible explanation for this is that sharing positive emotions does not happen as automatically. Still, it is important to note that displaying any emotion, even if it is positive, evokes more emotions than no or neutral emotions [49]. Therefore, when designing for empathy the differences in the extent different emotions evoke empathy is important to consider.

Next to the type of emotions, other situational factors also influence the experience of emotions. One of these factors is the way people express emotions through language and voice. How the target says something can influence to what extent the observer is perceiving emotions and is feeling empathetic [50]. In his study, Kraus [10] collected evidence that voice and sound seem to be the main communication channel for emotion and empathy. By relying only on voice the accuracy of recognizing emotions increases. Compared to expression through other modalities like facial expressions or through the multi-modal expression of emotions, vocal cues enhance the accuracy with which emotions are recognized [10]. Similarly, De Waele [51] argues that by hearing the emotions of a target, similar emotions can be evoked in the observer.

Interestingly, information about the emotional state of a speaker cannot only be communicated through words, but in most cases, non-semantic signals play a big role in this as well [52]. In their study, Regenbogen et al [53] looked at the effects of neutralizing different channels on the empathy evoked in the observer. They concluded that when the prosodic features of speech are neutral, it was significantly harder for people to feel empathy since the people had more trouble inferring the emotional states of the target. However, it should be noted that in the study the highest empathy was evoked when three different channels, specifically, facial expressions, prosody, and speech content, contained information about the emotional state. Accordingly, they concluded that all channels are important when evoking empathy [53]. Nonetheless, especially when training emotion recognition systems based on speech, prosodic information is highly important. Polzin and Waibel [54] developed a system to recognize emotions based on prosodic and acoustic information. When testing their system, a recognition rate of the emotions conveyed through speech similar to human performance was achieved [54]. Many similar examples can be found in the literature. Therefore, it can be concluded that non-lexical aspects of speech are highly important to display emotions.

In conclusion, when trying to evoke empathy, it is important that the target clearly displays the emotions. The more obvious the emotions are shown, the higher the likelihood that the observer experiences empathy. Also, people often experience more empathy in response to negative emotions compared to positive emotions. An important modality to communicating empathy-evoking emotions is through the voice. Interestingly, non-semantic signals are an important part of speech to evoke empathy. This gives reason to assume that musical utterances communicating emotions in HRI can be also successful in evoking empathy. However, to our current knowledge, no research exists to date regarding the extent to which the insights from human-human interaction transfer to HRI.

2.3.3 Empathy in Human-Robot Interaction

Generally, empathy is an important interpersonal concept fostering other interpersonal processes. Accordingly, in recent years, empathy became an important topic in the area of HRI. Systems are expected to interact with humans in a natural and social way by making use of social capabilities. This results, among others, in the need for empathy of and towards robots that interact with humans. Accordingly, there are two main streams of research regarding how empathy can be used in HRI. On the one hand, the robot can be the one showing empathy towards the user and their situation. In this case, the robot responds emotionally and empathically to the user [5]. For example, when agents show empathy towards their human interaction partner, the agent is perceived as more caring and supportive [55]. Also, empathetic robots are rated as more likable and engaging by the users [56]. These results show the importance of empathy in HRI.

However, the second stream of research regarding empathy in HRI is of greater importance for the current use case and research. Users can experience empathy towards a robot that consciously triggered this response. Here, the ultimate goal is to design for evoking empathy in the user [5]. The robot engages in the interaction in such a way, that the user responds with an emotional reaction and experiences empathy. Rosenthal-von der Pütten et al. [19] demonstrated in their study that when participants watched a video of a robot being treated badly, they felt greater empathetic concern for the robot. The participants felt greater negative emotions and less positive ones [19]. Therefore, it seems that already by observing a robot in an emotional situation, strong empathetic feelings can be evoked within the user. Additionally, in the FearNot! system, researchers were able to create virtual agents that evoked empathy within observers of a bullying situation. The participants experienced similar emotions in response to the observed situation and therefore, were empathetic toward the agents [20].

Nonetheless, one might question whether people can actually feel empathetic toward robots. Robots do not have any emotions and hence people can simply not appraise the situation in any way similar to them, as this process is not taking place within the robot. However, in 2.3.1 the importance of the appraisal of the situation was already discussed. Malinowska [57] was also concerned about the question of whether people can actually empathize with robots. Importantly, when humans interact with robots and hence enter a social relationship with them, they are treated as parts of one's social group. They become an individual of that group. Additionally, social robots are designed in such a way that they are easy to understand for people. Based on this, by the observation of robots, people can draw conclusions about what condition they are in [57]. This is similar to the use of curiosity of humans to understand the situation of another person when empathizing with them described before [45] [46]. Importantly, the conclusions people draw with regards to the robots are not based on what the robots experience, because they cannot do so, but rather on what state the robot communicates to be in [57]. Based on this, people can still appraise the situation of the robot in a similar way and hence experience empathy.

In general, empathy is an important factor in HRI. Previous studies have shown that virtual agents and robots are able to evoke empathy in users. However, to our knowledge, research about how robots can evoke empathy in their users is relatively limited, and not much research has been done on this, especially compared to the extensive research regarding how robots can show empathy.

2.4 Prosocial Behavior

Empathy is closely related to prosocial behavior which is important in the current use case for resolving issues of the robot. Prosocial behavior is any behavior with the purpose of purely benefiting others, for example by helping them resolve a problem. This behavior demands resources from and is not directly rewarding for the person engaging in the prosocial behavior. Even if this person receives rewards eventually, these are unexpected and unforeseen [58].

As described before, people that are more empathetic are more likely to engage in prosocial behavior. Empathy evokes altruistic motives to do something that is mainly beneficial to someone else instead of oneself [8]. Even when controlling for other variables like demographics and education, empathy remains a strong predictor of prosocial behavior [59]. Therefore, when one aims to design to evoke prosocial behaviors, one should also design to trigger empathy.

Importantly, prosocial behavior is a very broad concept and includes a wide range of different behaviors. Cooperative behaviors and actions are one form of prosocial behaviors that are positively influenced by empathy. Generally, there is a significant positive relationship between empathy and cooperation [60]. More concretely, compared to a condition of low empathy, people in high empathy conditions are more likely to make cooperative decisions. A feeling of empathy motivates people to act in a more cooperative way. Additionally, in situations in which detrimental outcomes might be expected, empathy-motivated cooperation is highly important to reduce and even eliminate these effects [61]. Therefore, empathy seems to be an important factor in situations where interactions take place and people have to work together.

However, more interestingly for the current use case, helping others, humans as well as other kinds of species, is another form of prosocial behavior. Generally, it can be said that empathy is positively correlated with the intention to help someone. People that experience higher levels of empathy are also more willing to help that person [8]. Accordingly, empathy should be elicited within a person when one would like to get help from this person.

Concluding, prosocial behavior is any behavior benefiting solely someone or something else. In previous research, prosocial behavior and sub-constructs have been shown to be influenced by empathy. However, most of these insights are from research regarding human-human interaction and it remains unclear whether this relationship can be also found in HRI. Also, the question arises, whether we can elicit prosocial behavior through communicating intentions and emotions and evoking empathy through musical utterances.

2.5 Research Objectives

Based on these insights from previous research, we are able to formulate concrete research questions and hypotheses for our overall research question. We test our hypothesis regarding musical utterances in comparison to no sounds and beeping sounds. We refer to these in the hypothesis as other sound conditions. We included beeping sounds as a comparison group, as these are another, possibly simpler, form of non-semantic communication. Beeping sounds can be regarded as non-linguistic utterances. It has already been shown that these can communicate emotions in HRI. Therefore, we are interested in whether musical utterances, so more complex sounds, even increase this effect. Additionally, beeping sounds were excluded to ensure that any effects observed compared to the no sound condition are not only to the attention-grabbing nature of novel sounds [62].

The first research question relates to the legibility of robot behavior. We are interested in whether musical utterances are a way to communicate the intentions of a robot and therefore they become more legible. Accordingly, the first research question is to what extent the expression of intentions is legible when robots communicate by using musical utterances. Other types of SFU have been shown to be well-understood by users of robots. Often, these are easily understood by users due to their simplicity. In studies about legibility, non-verbal behavior is important to enhance this understanding and interpretation of the users. Therefore, our first hypothesis is:

H1: When using musical utterances, the communication of the robot is more legible compared to other sound conditions.

Secondly, we are interested in the emotions evoked by empathy. The second research question is to what extent musical utterances are able to communicate emotions that are closely related to empathy. Although, first studies already show that emotions are perceived by users of robots who communicate these through musical sounds. Nonetheless, in those studies, the aim is to communicate emotions only. In our study, however, we do not only communicate emotions but also try to communicate intentions. Therefore, the question arises whether the emotions can be still well communicated. Research shows that music can convey emotions well (see for example [33] [34]). The manipulation of musical features closely relates to which emotion is perceived by the listener. Also, recent human-robot studies show that people perceive emotions in musical sounds made by robots. Hence, the second hypothesis reads as follows:

H2: When using music utterances, emotions are communicated more clearly com-

pared to other sound conditions.

As the communication of emotions relates closely to observers experiencing empathy, we are considering whether this is applicable in the current use case. Our research tests whether the communication of emotions through musical utterances is able to evoke empathy. Therefore, building on the second research question and hypothesis, the third research question concerns to what extent people experience cognitive empathy when emotions are communicated by musical utterances and to what extent people experience affective empathy when emotions are communicated by musical utterances. Research has shown that people can feel empathy towards robots, even though robots do not experience emotions themselves. However, to our knowledge, no research has yet been done on whether emotions in musical utterances can trigger empathy. Generally, if people perceive the emotions of others in human-human interaction, it is likely that they feel empathetic. Empathy here is conceptualized as the interplay between cognitive and affective empathy. Research has shown that only through this interplay empathy can be experienced. Therefore, the third hypothesis is divided into two sub-hypotheses:

H3.1: People experience more cognitive empathy when hearing musical utterances interacting with a robot compared to other sound conditions.

H3.2: People experience more affective empathy when hearing musical utterances interacting with a robot compared to other sound conditions.

Eventually, when people experience empathy, it triggers other interpersonal processes in human-human interaction. One of these processes is prosocial behavior. Therefore, lastly, we are concerned with the engagement of people in prosocial behavior when they hear musical utterances as means of communication. Accordingly, the fourth research question is to what extent people intend to engage in prosocial behavior when hearing musical utterances. When feeling empathetic, people are also more likely to engage in prosocial behavior as empathy evokes altruistic motives. Several of the sub-constructs, as well as prosocial behavior itself, are positively influenced by empathy in human-human interaction. Building on the previous hypothesis, if people are more empathetic in response to musical utterances, we also expect them to be more likely to intend to engage in prosocial behavior. Hence, the last hypothesis is the following:

H4: When communicating through musical utterances, people intend to engage more in prosocial behavior compared to other sound conditions.

Chapter 3

Sound Design

To evaluate the use of musical utterances as means of communication when the robot has a problem, we designed sounds with sound designers for such scenarios. In the following, we will first describe the setup and outline of the sound design session. Then we will describe the sounds as an outcome of the session and finally, we will outline some design considerations that were raised during the design session.

3.1 Sound Design Sessions

The aim of the design sessions was to create sounds that fit selected scenarios in the context of the current research. Therefore, we were aiming for sounds that conveyed specific intentions and emotions of the different scenarios. We wanted to get sounds that fit the appearance of the robot and the environment of a hospital.

3.1.1 Sound Designers

Two sound designers were recruited through convenience. These sound designers were both male and have more than six years of experience with amateur music production within the electronic music production sound space. Both started their music journey by learning how to play instruments. Their skills with regard to music production are mostly self-taught. One of the designers also has experience in sound design for installation and functional interactions through various courses. The other one also gained knowledge and skills by doing research on music, by for example interviewing producers about their experience. The main capabilities of the sound designers lie within the digital synthesis of sound, meaning the creation of sound by a computer. The sound designers mainly produce music in the genre of dance music, house, techno, ambient, and funk.

3.1.2 Equipment and Materials

For the sound design session, we created scenario outlines (see Appendix A). The selection of the scenarios was partly based on possible problems the Harmony project identified. Additionally, we choose scenarios in which the actions to resolve the problem of the robot varied to test different kinds of contexts. The scenarios were the people blocking the robot's way, a door being closed on the way of the robot, a cable on the floor on the way of the robot, and the robot losing navigation. They included an introduction in which the general setting was described. This was followed by a warm-up scenario. Here, the situation of a robot entering an elevator with people was described. Importantly this scenario was intended to be similar to the main scenarios but at the same time also different enough to not influence the sound design for the main scenarios. A suitable scenario seemed to be one where the robot was encountering people instead of having a problem. Finally, the four main scenarios followed. All scenarios were written as design fiction. Design fictions are artifacts that describe constructs of technology that do not exist yet [63]. They play in the somewhat near future and help envision the interaction and experience with these technologies. Importantly, it does not concern what is currently technologically possible but rather opens the space for discussion and exploration [63]. Through the design fiction, we were able to create scenarios that the sound designers could well imagine themselves in and therefore relate to. At the end of the design fiction, it was summarized with one sentence to highlight the most important aspects. An example of such a sentence is 'What sound does the robot make to get the intention that **the people need to move out of the way** across in an **emotional** way so that it might trigger **empathy**?'.

To design the sounds, the sound designers used the digital audio workstation Ableton Live together with a set of virtual studio technology (VST) plugins. A 49-key midi-controller was used as input to control the different VST plugins. The VST plugins used were JUP-8 V by Arturia, which is a synthesizer, and E-Piano by Ableton.

3.1.3 Session Outline

In total, we had two sessions. In the first one, an introduction session, we described the Harmony project to the sound designers to give a general idea of the project. Then, we explained the goals of the current research to the sound designers. Following that, the sound designers watched a short video showing the prototype of the robot from the project to get an idea of the behavior of the robot. The robot has a head with white eyes and lights on the sides of the head. It does not have any arms or legs. At the lower part of the body, it has again lights. All lights can light up in different colors. The robot moves on wheels. The video was from a TV show which filmed the robot in a hospital environment. Additionally, the sound designers saw two pictures of the robot (see Figure 3.1). These pictures were shown on a screen during the entire sound design session. The aim of this was that the sound designers had a good idea of what the robot looked and behaved like to design sounds that were in line with this. Then we explained the aim and the goal of the design session. The sound designers were instructed to design sounds for the four scenarios. The main aim of the sounds is to convey the intention of the robot while being emotional. The decision on the specific types of emotions was up to the interpretation and creativity of the sound designers. Additionally, we also asked the sound designers to design for a diverse group in a hospital, to be as short as possible without getting the characteristic of more technical, computer-like sounds, and to have musical qualities. Again, the interpretations of the latter two aspects were left to the sound designers. Also, we instructed the sound designers to explain the intention and rationale behind the sounds they designed. Then, the designers received the scenarios for the sound design. They had some time to read through all scenarios once to get some idea about the setting and general outline. Once the sound designers had a good understanding of the intentions and goals of the sounds we ended the introduction session.

In the second session, we started by recapping the most important information regarding the sound design. Initially, we did a short warm-up with the according scenario. The sound designers had time to get used to the instructions in this way. Afterward, we started with the first scenario. The sound designers read through the scenario description again more thoroughly. Then, they started to think about the sounds, exchanged ideas, and experimented with different, sounds melodies and pitches. While doing so, they voiced their thoughts and ideas behind what they were doing. When the sound designers felt like they were done or would not advance with the sound at that moment, they stopped working on that sound and continued to the next scenario. This procedure was repeated for all scenarios. When the sound designers felt confident about all sounds designed, the session ended.

3.2 Designed Sounds

Overall, the sound designers explained that they were aiming to design sound with a relatively higher pitch to match the robot they perceived as cute. Through this, they tried to match the appearance of the robot and especially the eyes. For each scenario, they described that they structured the sound in such a way that it first grabs the attention of the people around the robot and then conveys the intentions and emotions of the robot. The result of the sound design session was therefore func-

Figure 3.1: Photos of the Harmony robot shown to the sound designers.

tional music[1], meaning music with a specified aim and message. This was based on melodies, chords, and rhythms that are closely connected to classical music theory. This helped to create a warmer and more suitable feeling in the utterances. The resulting sounds all lasted between three and four seconds.

3.2.1 People Blocking the Way of The Robot

For the first scenario, people blocking the way of the robot, the sound designers had the idea of making a sweeping sound to imitate the movement the sound designers expected as a common response of the people to do, so to step aside. By adding chords to the sounds they aimed to give the sound a nicer and more friendly touch. In the first part, the sound communicates that something is wrong with a sound similar to soft beeping. At the end of the sound, staccato piano sounds aim to convey a nice but determined sound to move away (see Figure 3.2). The idea of the sound designers was to keep the sound simple and friendly as the robot wanted something from the people while simultaneously conveying some tension as the sound has a relatively fast pace to convey urgency.

[1] The designed sounds can be found here.

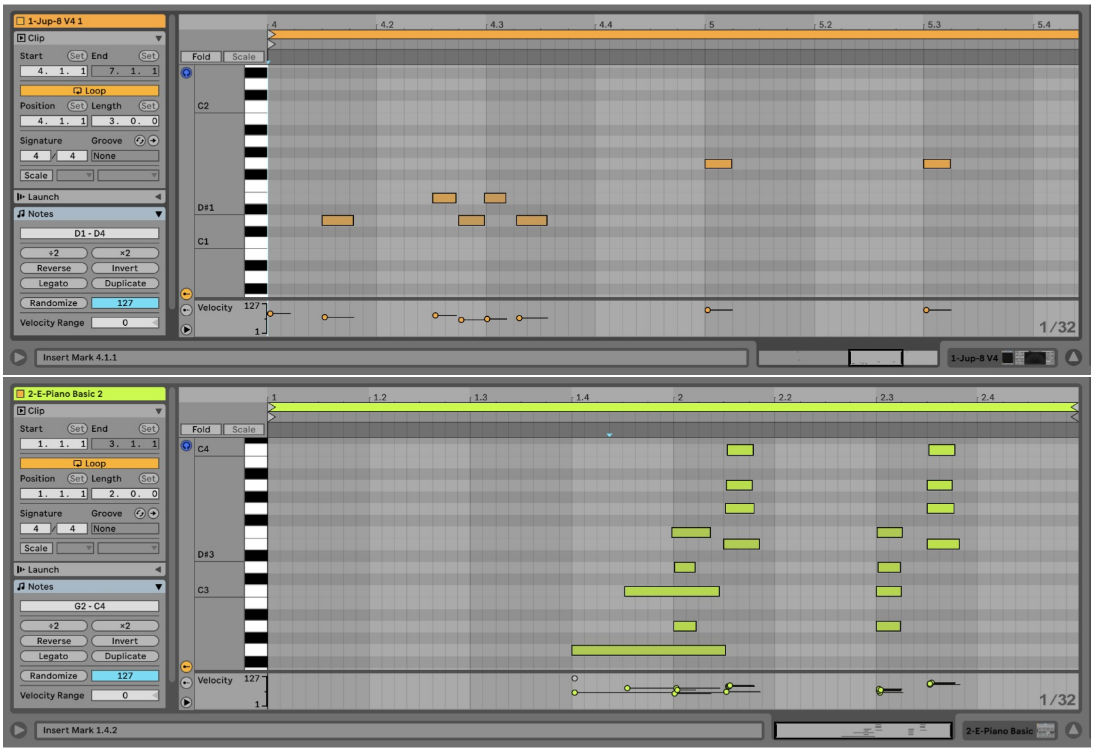


Figure 3.2: Musical scores of the sound for the scenario where people are blocking the way of the robot. The orange scores are for the Jup-8 V plugin and the green sounds are from the E-Piano plugin.

3.2.2 Door Being Closed

The idea behind the sound for the second scenario, a door on the way of the robot being closed, was to be more alarming. The robot has to use the moments when people are passing by to not stay and wait there longer. Therefore, the designers intended to make the beginning of the sound more attention-grabbing. This alarming part was enriched by adding more musical sounds. Additionally, the sound designers intended to imitate a knocking sound as if the robot is knocking on the door (see Figure 3.3). This part was also determined to show some level of distress.

3.2.3 Cable on the Ground

For the third scenario a able being in the way, the sound designers tried to resemble some sort of stumbling or tripping sound. At the same time, they also wanted it to sound like a few steps were taken. Therefore, the sound started with two short different tones playing alternately. This is followed by a sequence of tones going down and being played more legato (see Figure 3.4). The latter part was also intended to make the sequence sound more dramatic to enhance the problem being conveyed.

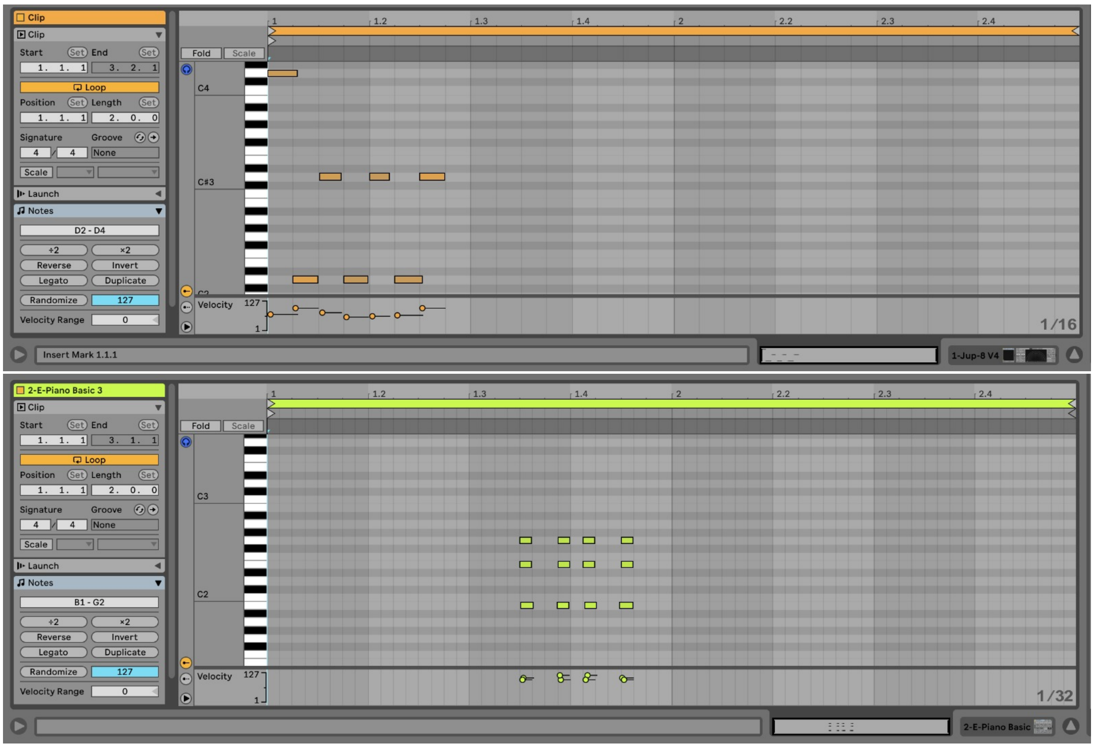


Figure 3.3: Musical scores of the sound for the scenario where a door on the way of the robot is blocked. The orange scores are for the Jup-8 V plugin and the green sounds are from the E-Piano plug-in.

3.2.4 Robot Losing Navigation

For the fourth and last scenario, the main idea of the sound designers was to have some kind of sensing sound as a basis. They wanted to convey that the robot is scanning its environment and trying to make meaning out of it, which is unsuccessful because it does not know where it is. At the same time, they wanted to convey that the robot is sad because it is lost. Hence, they came up with a relatively slow sound, again starting with alternating tones and then going down at the end (see Figure 3.5).

3.3 Design Considerations

Overall, the sound designers were able to come up with sounds for all scenarios. They easily came up with ideas of what sound would fit the scenario and what features the sound should have. Putting these ideas into practice was more difficult. They needed some experimentation and tryouts to match their ideas with what they created. However, they also mentioned that this is the normal process of sound design. Nonetheless, as they had a unique assignment and therefore, this difficulty

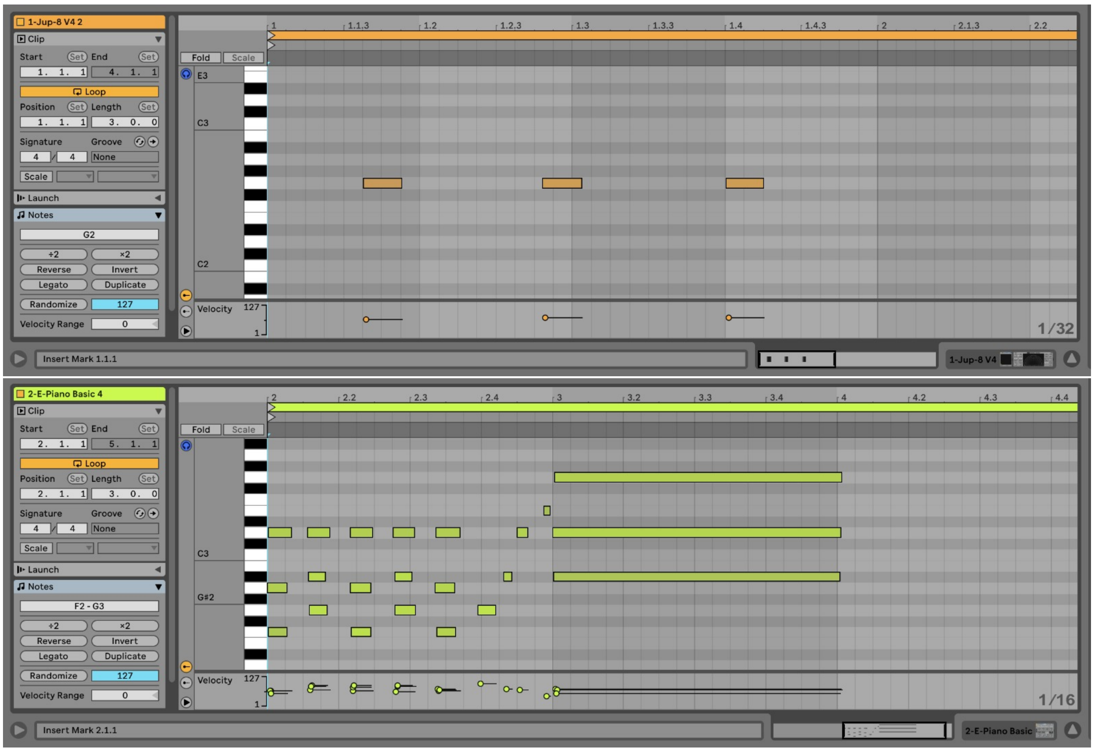


Figure 3.4: Musical scores of the sound for the scenario where a cable is in the way of the robot. The orange scores are for the Jup-8 V plugin and the green sounds are from the E-Piano plugin.

might have been even bigger. Designing sounds for the fourth scenario (the robot being lost) was the most difficult. The sound designers had trouble coming up with an idea for a sound and how to put it into practice. They mentioned that the difficulty was that the error was not easily externally visible to which they could refer to with the sound. Therefore, it was harder for them to design the sound and they also were not fully satisfied with it eventually. Still, they also did not have an idea of how to improve it.

However, there were also some other considerations mentioned by the sound designers that would have influenced the sounds. First, they mentioned that it would have helped them to see the videos of the scenarios themselves. With the design fictions of the scenarios, they already could imagine well what the scenarios are looking like. Nonetheless, the sound designers mentioned that with the actual videos the fit between the sound and scenarios possibly could have been even better. They would have seen the timing in the videos and how the different parts of the context interplay with each other. This could have given the designers valuable information to make the sounds fit closely with the scenarios.

Another aspect raised by the sound designers was the diversity of music. In the sound design session, we almost completely relied on the creativity of the sound de-

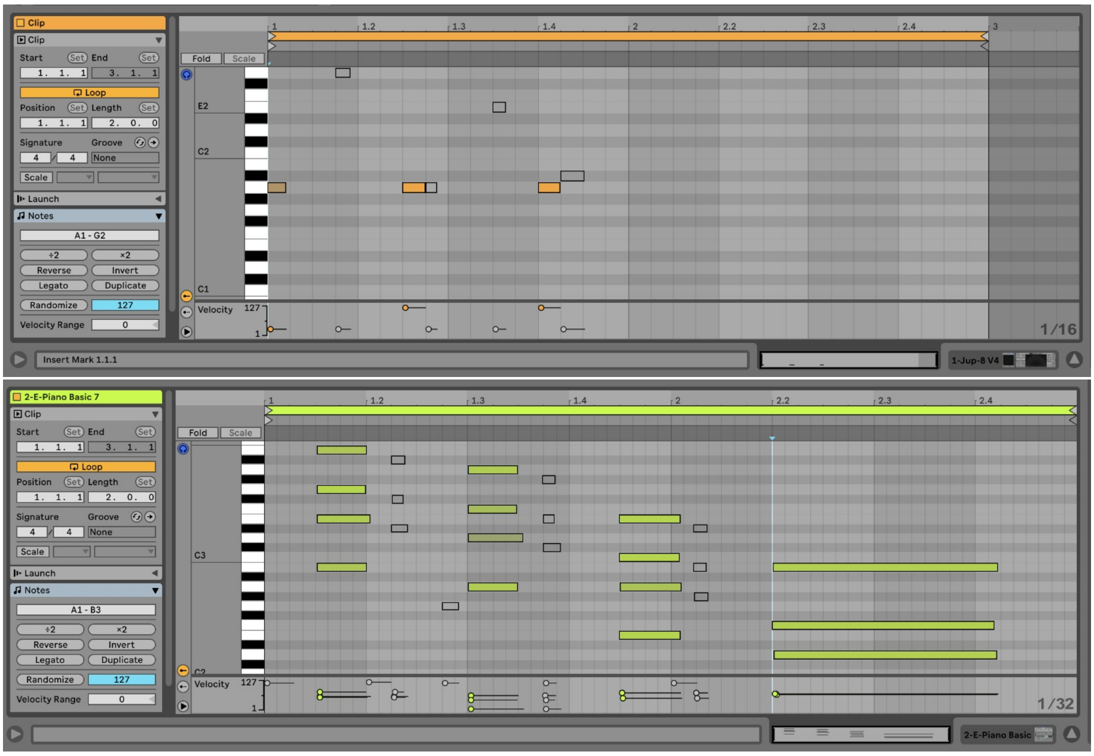


Figure 3.5: Musical scores of the sound for when the robot has lost its navigation. The orange scores are for the Jup-8 V plugin and the green sounds are from the E-Piano plugin.

signers and fully on their imagination and experience in the context of the hospitals. Therefore, they had complete freedom. Yet, sound design is much more complex. Music is very diverse and often you can achieve the same result in many different ways. Hence, finding the right way and studying the factors influencing this could be helpful. For example, it could have been helpful to get insights from people in a hospital regarding what kind of sounds they want to have added to their already noisy environment. This would have given the sound designers close insights into what to take into account when designing. However, this was out of the scope of the current study.

Lastly, the sound designers mentioned that they had to make a trade-off between functionality and musicality. If they would have focused fully on either of the two, the other aspect would have moved into the background and might not have been as much resembled as much. Therefore, they created sounds that conveyed the intentions as much as possible while still being considered musical by them. This also means that they might not be the most functional sounds and that a fully functional sound would have been a lot less musical and vice versa. The sound designers tried to find a middle way as much as possible.

3.4 Conclusion

Overall, the sound design went well and the sound designers were able to create sounds for the scenarios. They were confident in the match of the sounds with the scenarios, the robot, and the other design requirements. Hence, we were able to obtain musical utterances for the robot. However, in hindsight, there are some design considerations that could have not been taken into account in the sessions. These considerations possibly would have improved the fit of the sounds even more. However, they were out of the scope of the current research. Therefore, we have a first set of musical utterances that can be tested. However, for future design of sounds the considerations from section 3.3 should be taken into account.

Chapter 4

Methods

To evaluate the sounds designed in the sound design session, we conducted an online video study. We wanted to test whether musical utterances were able to improve the legibility of a situation, evoke more empathy-evoking emotions and empathy itself, and increase the intentions to engage in prosocial behaviors. In this Chapter, we describe the study.

4.1 Participants

The participants were recruited through the crowd-sourcing platform Prolific. All participants received 3$ for their participation. The inclusion criteria for the study were that the participants had to be from the United States of America, do not have any hearing or cognitive impairments, have completed 100 or more studies on Prolific, and have an approval rate of 95% or higher. In total, 199 participants participated in the study. From these, four participants were excluded because their completion time was very long or very short, they did not answer one or more of the attention checks correctly and their answers also gave reason to assume they did not pay close attention. Another seven participants were excluded because they did not pass several attention checks, which gives reason to assume they did not pay close attention to the videos which again was also indicated by their answers. This resulted in a total of 187 respondents (87 female, 99 male, 1 non-binary) that were included in the data analysis. The mean age was 36.76 years (SD = 12.23). The participants were also asked about their experience with robots (see Figure 4.1).

The study was reviewed and approved by the ethics committee of the faculty of Electrical Engineering, Mathematics, and Computer Science of the University of Twente and registered under the reference number 230072.

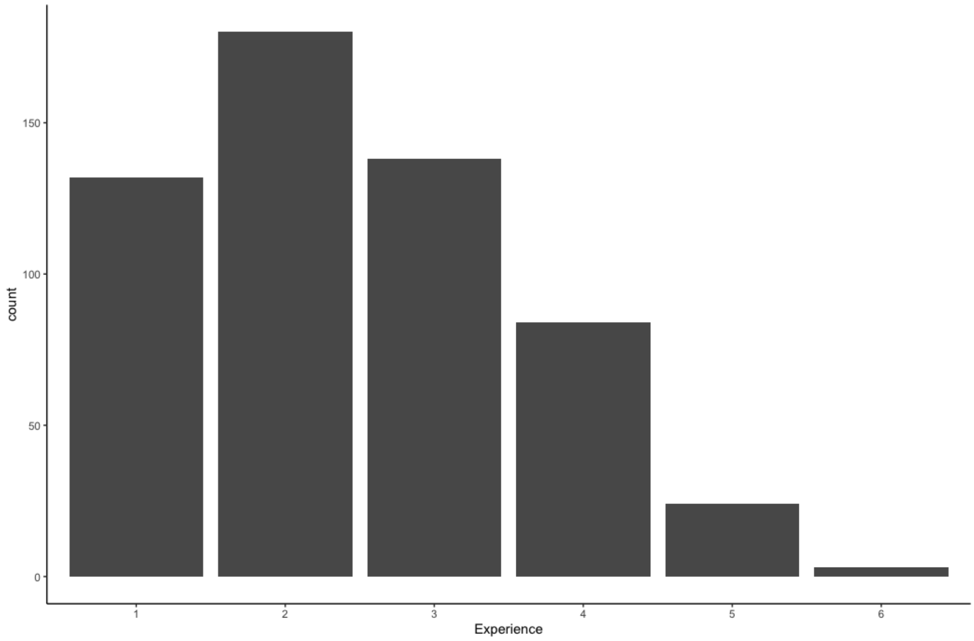


Figure 4.1: Frequencies of ratings of experience with robots by participants.

4.2 Stimuli

In the study, videos of the robot in different scenarios were used (see Appendix B for the scripts). The scenarios in the videos were the same as the ones used for the sound design. Hence, in these videos, the robot was stuck and could not proceed on its journey without the help of the people around the robot[1]. Due to scheduling difficulties, the robot developed by the Harmony project was not available in the time frame of recording the videos. As the Kuka ido robot looks similar to the robot of the project this one was used in the videos. See Figure 4.2 for screenshots of the videos and the composition of the robot. We steered the robot with a remote control to drive through the hallway. As described earlier, musical utterances and beeping sounds were used in two conditions. The musical utterances were taken from the sound design session described in Section 3. In the sound design session, one sound was designed for each scenario, hence these were used. The beeping sound was created by one of the sound designers from the sound design session. The designer was instructed to design a repeating beeping sound with the intention of grabbing attention. Additionally, the sound should have the same length as the musical utterances as well as a similar pitch as the musical utterances. These sounds were played and recorded separately in the hallway of the videos and added afterward to the videos. This was done to use the exact same videos for all conditions to exclude any possible influence by differences in the videos. The videos were recorded

[1]The recorded videos can be found here.

with the Panasonic Lumix DC G-90 camera and a standard 12-60mm kit lens. The sound was recorded by attaching a Rode Video mic Go through an AUX cable to the camera.

Figure 4.2: Screenshots of the videos used in the questionnaire.

4.3 Design

We conducted a video study about HRI with regard to the influence of different sounds. The study was designed as a 1 by 3 between-participants design. The independent variable was the sound condition which has three levels. The variable was a between-participant variable where each participant was randomly assigned to one condition. The first level was the no sound condition. Participants in this condition saw the videos without the robot making any sounds. In the beeping sound condition, the robot made beeping sounds in the video. We included this condition to see whether the complexity of music has an added value compared to a simple beeping sound. As a third condition, the robot played the musical utterances designed previously. Initially, all four different scenarios were included in the study.

However, during the pilot run of the study, all participants had difficulty with the scenario regarding the robot losing its navigation. As this scenario already showed some difficulties during the sound design, we decided to not make use of this scenario. Hence, the study only concerned scenarios where there is a visible external cause for the problem of the robot. The random effect was hence measured on three different levels and each participant saw the robot in all scenarios. Accordingly, this is a within-participant variable.

4.4 Procedure

SurveyMonkey was used as a platform for the questionnaire. Participants received a link through Prolific to access the questionnaire on SurveyMonkey. Participants received an introduction to the questionnaire and the use case at the beginning of the questionnaire. It was explained that the videos take place in a hospital and their tasks were described. After this, the participant answered the question regarding their experience with robots. To ensure that the participants had their sound turned on, a test video was shown to them. This was a video of the robot making similar sounds to the sounds designed previously. All participants saw the same video with the same sound, regardless of the condition they were in. The participants answered an open question about what the sound sounded like to ensure they had their sound turned on. Then the participants watched the first video. Initially, the participants were asked a question to check whether they paid attention to the video until the end. Then the questions regarding legibility, emotion recognition, empathy, and prosocial behavior were asked in this order. The statements about empathy were shown to the participants in a random order. This procedure was repeated for the remaining two videos. At the end of the questionnaire, the participants were asked about their age and gender. Prior to filling out the questionnaire, the participants received an information brochure and had to consent to participate. After the background questions, the participants were de-briefed and were again asked to give their consent to ensure they still consent after having all information after they received the full information.

4.5 Measurements

All measurements were implemented into a questionnaire (see Appendix H).

4.5.1 Legibility

The measure of legibility was done similarly to Kim and Follmer [64]. The participants had to answer the open question "What do you think the robot is trying to communicate?" and rate their confidence in their answer on a 7-point Likert scale. Here, 1 meant that they were not confident at all and 7 meant that they were completely confident. The open question was coded based on the level of accuracy with which the participants described their answers. We included three levels of accuracy in the coding scheme. No description of the problem (0) meant that the participants did not recognize a problem or recognized a wrong problem that was not actually a problem. The second level, a general description of the problem (1) indicated that the participants recognized that there was something wrong with the robot but did not describe the problem in further detail. Lastly, a detailed description of the problem (2) meant that the participant described the problem that the robot actually had accurately(see Appendix C for the coding scheme). A second rater coded 40% of the answers. The inter-rater reliability was indicating substantial agreement (κ = .77). However, when looking at the confusion matrix (see Table 4.1), a lot of disagreement relative to the total number could be seen for when to rate something as 1 and when as 2. When looking at the cases where this disagreement occurred, the disagreement was very diverse and it could not be accounted for one specific reason. This gives reason to assume that the differentiation might be hard to do or there might not even exist such a difference. Therefore, we decided to merge categories 1 and 2 into one category. Therefore, it was a matter of whether the communication was legible.

Table 4.1: Confusion matrix of the ratings for legibility.

	Rater 2			
Rater 1	0	1	2	Total
0	15	1	0	16
1	0	6	3	9
2	1	5	92	98
Total	16	12	95	123

4.5.2 Emotion Recognition

To measure the recognition of emotions, we asked the open question "What emotion is the robot displaying? (Think about emotions like happy, sad, disgusted, neutral, angry)". This question was coded as neutral (0), positive (1), and negative (2) emotions (see Appendix D). We decided to code the perceived emotions in this way because for empathy negative emotions are the most important. As described in Section 2.3.2, people are most empathetic in response to negative emotions. Hence, in our study, these are the most desirable emotions. A second rater coded 40% of the answers. The inter-rater reliability was strong (κ = .88). Also the confusion matrix (see Table 4.2) did not give any reason for concern. In addition, the participants indicated on a 7-point Likert scale how strongly they perceived the emotion. Here, 1 meant the emotions were not at all strong and 7 meant they were very strong.

Table 4.2: Confusion matrix of the ratings for emotions.

	Rater 2			
Rater 1	-1	0	1	Total
-1	48	5	1	54
0	1	60	1	62
1	0	0	7	7
Total	49	65	9	123

4.5.3 Empathy

Empathy was measured by rating a set of statements related to situational empathy on a 7-point Likert scale. Here, 1 meant the statement was not true at all and 7 meant it was completely true. The statements were inspired by the statements of the Questionnaire of Cognitive and Affective Empathy [65]. This questionnaire was designed to measure the levels of dispositional cognitive and affective empathy. However, in our study, we were measuring situational empathy. Accordingly, we selected statements that were applicable to the use case content-wise and reformulated them to concern situational empathy. In total, we included five statements measuring cognitive empathy and four statements measuring affective empathy. An example of a statement for cognitive empathy was 'It was easy to imagine how the robot was feeling' and a statement for affective empathy was 'The robot influenced how I am feeling' (see Appendix E). We ran factor analyses on the statements for each scenario separately. Overall, these confirmed that there are two underlying factors. However, the factor analyses also pointed out that the statements 'I stayed emotionally detached from the situation' and 'I considered the feelings of the robot' are problematic. The former statement loaded low on both factors for all scenarios. The latter loaded high on both factors and depending on the scenario also loaded higher on one or the other factor (see Appendix F). Looking at the content of the statements, they also seemed more difficult. The former problematic statement was the only statement reversed coded. The second problematic statement was about the feelings of the robot. However, many participants mentioned in the open question about the emotions that robots do not have emotions. Therefore, this question could have been confusing to them. For these reasons, the two statements were excluded from the analysis. The combined factor analysis for all scenarios together excluding the problematic statements showed excellent reliability. For cognitive empathy, Cronbach's alpha was .93 and for affective empathy, it was .92. Both factors together explain 77% of the variance in our data (see Appendix F for the factor loadings).

4.5.4 Prosocial Behavior

To measure whether the participants intended to engage in prosocial behavior they were asked the open question "What would you do if you were a person in the situation?". The answers were coded as no prosocial behavior (0), weak prosocial behavior (1), and strong prosocial behavior (2) (see Appendix G). No prosocial behavior meant that the participant indicated that they would not do something to try to resolve the situation of the robot. Weak prosocial behavior was behavior related to resolving the problem but clearly not resolving the problem. Lastly, strong prosocial behavior meant that the participants would do anything they could do to resolve the situation. A second rater coded 40$ of the answers. The inter-rater reliability was indicating a strong reliability (κ = .83). In the confusion matrix (see Table 4.3) it was again apparent that there was some disagreement between weak and strong prosocial behavior. When looking at the answers that caused disagreement, no pattern of disagreement could be observed. Since the disagreement was proportionally not too big and Cohen's kappa also indicated strong reliability, it was decided to leave the coding as it is.

Table 4.3: Confusion matrix of the ratings for prosocial behavior.

	Rater 2			
Rater 1	0	1	2	Total
0	21	1	1	23
1	4	26	5	35
2	0	2	63	65
Total	25	29	69	123

Chapter 5

Results

For the analysis of the outcomes of the study, it is important to note that we used the musical utterance condition as a reference condition. Although this practice is rather unusual, it enables us to see the impact musical utterances have compared to other sound conditions which was what we were most interested in.

5.1 Legibility

In Figure 5.1, the distribution of the participants depending on the sound can be seen. An ordinal logistic mixed model was run to test whether the sound condition had an influence on whether the communication of the robot was legible or not. The scenarios were included as a random effect. Age and experience with robots were included as covariates separately to check whether they had a significant influence as well and improve the model. However, the model fit was not improved by age ($\chi 2$ (1) = 2.93 p = .087) or experience ($\chi 2$ (1) = 0.14, p = .705) and both also did not have a significant influence (see Appendix I). Therefore, they were not included in the model. Analysis showed that the sound condition did not have a significant effect on the legibility of the communication of the robot (see Table 5.1).

Table 5.1: Ordinal Logistic Mixed Model for Legibility.

Coefficients	B	Std. Error	t	Sig.
No sound	-0.30	0.27	-1.13	.260
Beeping	0.16	0.29	0.54	.590

Note. The musical utterance condition was the reference category.

Legibility was also measured as the confidence the participants had in the answer they gave to the open questions. Figure 5.2 shows the distribution of the level of confidence in the participants' answers regarding the interpretation of the robot communication. A linear mixed model was run to test whether the sound had an

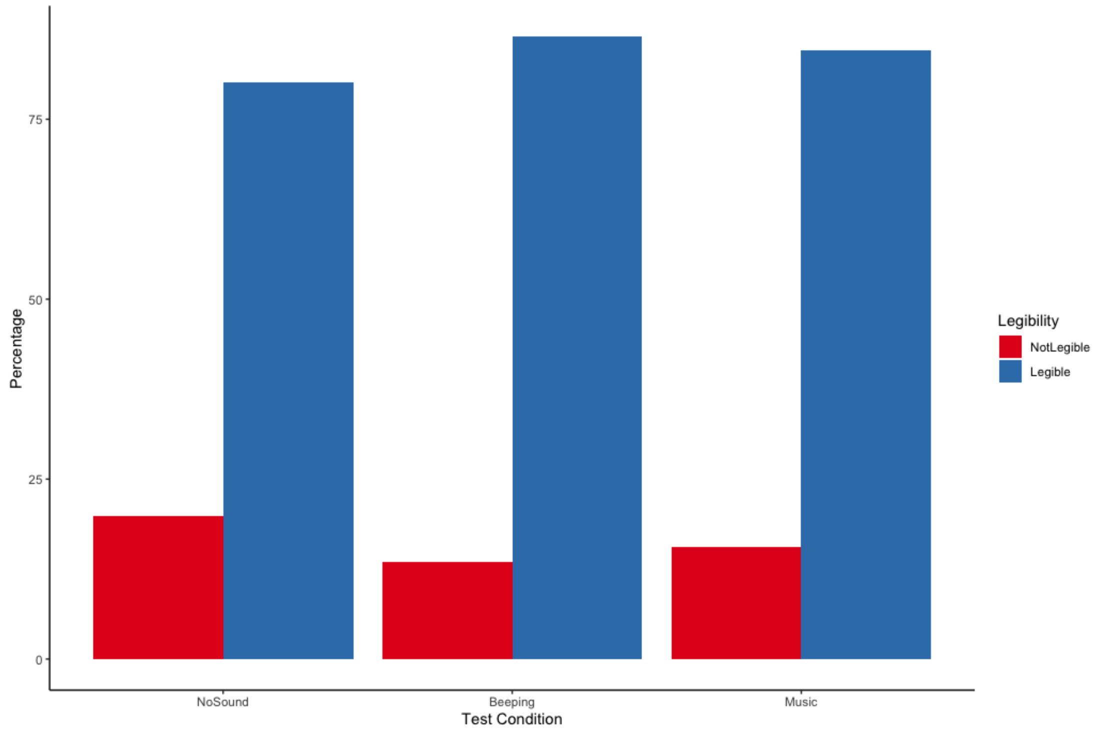


Figure 5.1: Bar chart showing the distribution of legibility per condition.

influence on confidence with the scenarios as a random effect. Again, we also included age, gender, and experience with robots as covariates separately. Age had a significant influence on confidence and significantly improved the model fit ($\chi2$ (1) = 16.44, $p \leq .001$). Experience with robots was found to have a significant influence on confidence and also improve the model fit ($\chi2$ (1) = 8.00, $p \leq .001$). Therefore, the covariates were included together in the model. Also then, both had a significant influence on confidence and improved the model fit again ($\chi2$ (1) = 12.35, $p \leq .001$) (see Appendix I). Therefore, age and experience with robots were included as covariates in the model. To test the fit of the model with the data, we checked the assumptions of normality, equal variance, and linearity. All of them were approximately met (see Appendix J) and hence a linear mixed model was appropriate. The analysis showed that compared to the beeping sound condition, participants were significantly less confident in their answers to the open question in the condition of the musical utterance. However, compared to the no sound condition, the participants were not significantly more confident in the musical utterance condition. Additionally, the more experienced with robots people were the more confident they were in their interpretation and the older people were the more confident they were as well. There was a significant positive relationship between age and experience and confidence in the interpretation of the communication (see Table 5.2). The mean confidence scores were 4.77 (SD = 1.66) for the no sound condition, 5.62 (SD

= 1.50) for the beeping sound condition, and 5.13 (SD = 1.61) for the musical sound condition.

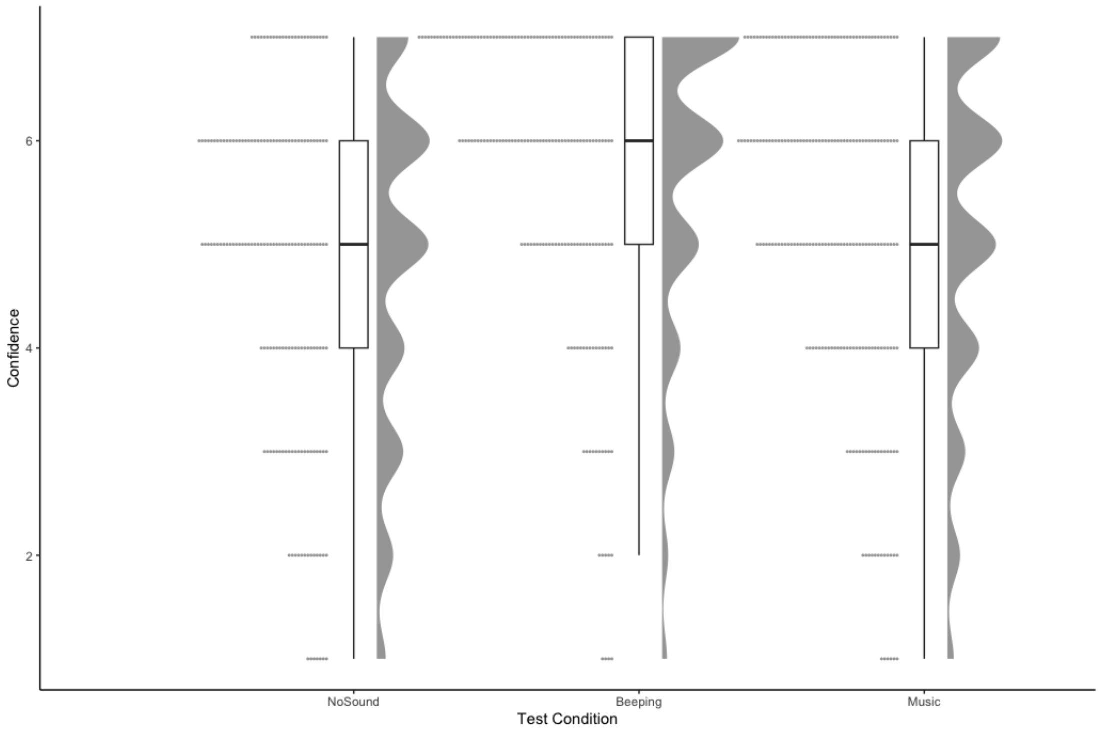


Figure 5.2: Raincloud plot showing the distribution of the confidence in the answer about legibility per condition.

Table 5.2: Linear Mixed Model for Confidence.

Coefficients	B	Std. Error	t	Sig.
Intercept	3.68	0.31	12.02	.000
No sound	-0.27	0.16	-1.69	.091
Beeping	0.53	0.16	3.35	.001
Experience	0.20	0.06	3.53	.000
Age	0.02	0.01	4.60	.000

Note. The musical utterance condition was the reference category.

5.2 Perceived Emotion

Figure 5.3 shows the distribution of the perceived emotion per condition. We ran an ordinal logistic mixed model to test the influence of sound on perceived emotion. Here as well, age and experience with robots were included separately as covariates. Age had a significant influence on perceived emotions and significantly improved the model fit ($\chi2$(1) = 4.09, *p* = .043). Also, experience with robots had a significant influence on the perceived emotion and it also improved the fit of the model ($\chi2$(1) = 4.38, *p* =.036). Therefore, we also included them together as covariates in the model. In this case, the influences of age and experience were no longer significant and the model fit was also not better ($\chi2$(1) = 3.02, *p* = .082, see Appendix K). Hence, we considered two models with age and experience separately. In both models, a significant difference between the beeping sound and musical utterance condition was found. For the musical utterance condition, there was a higher log odd ratio that the people are in a higher category, hence perceiving a level of emotion that elicits more empathy. However, no significant differences were seen between the no sound and musical utterance conditions. Experience with robots and age had a significant influence on perceived emotion. The older people were, the lower the log odds to perceive emotions that enhance empathy. Contrary to that, the more experienced people were with robots the higher the likelihood that they perceive emotions in the robot communication that were related to higher empathy (see Table 5.3).

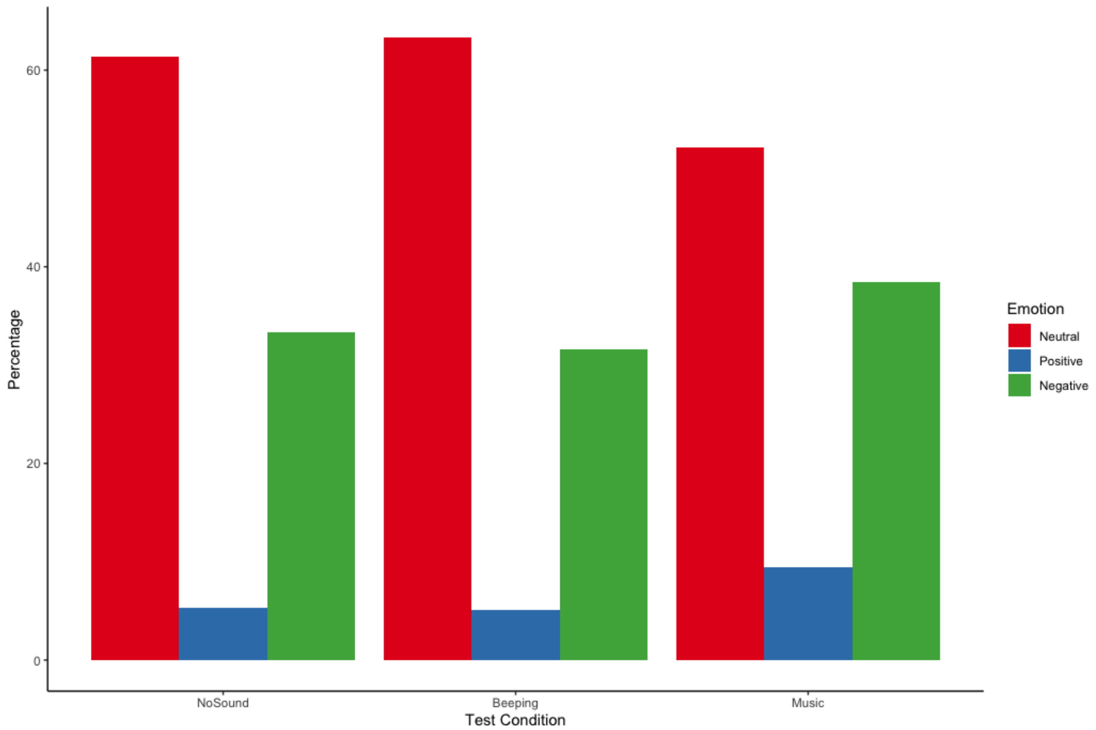


Figure 5.3: Bar chart showing the distribution of the perceived emotion per condition.

In Figure 5.4, the distribution of perceived emotion strength for every sound condition can be seen. By using a linear mixed model, we tested the influence of sound on the perceived strength of the identified emotion. Again, we included age, gender, and experience with robots as covariates separately. We did not find any significant

Table 5.3: Ordinal Logistic Mixed Model for Perceived Emotion.

Coefficients	B	Std. Error	t	Sig.
Model with age as covariate				
No sound	-0.36	0.21	-1.78	.079
Beeping	-0.43	0.20	-2.11	.035
Age	-0.01	0.01	-2.00	.045
Model with experience with robots as covariate				
No sound	-0.30	0.21	-1.48	.138
Beeping	-0.40	0.20	-1.78	.048
Experience	0.02	0.01	2.56	.037

Note. The musical utterance condition was the reference category.

influence of age and experience with robots on the strength of emotion. However, being non-binary had a significant influence on the perceived strength of the emotion. Nonetheless, there was only one participant identifying as non-binary. Additionally, the model including gender did not show an improved fit of the model ($\chi 2$(1) = 5.61, *p* =.061). Including age in the model did significantly improve it ($\chi 2$(1) = 5.34, *p* =.021). Therefore, we decided to include age in the model and not gender (see Appendix K). The fit of the model with the data was checked by looking at the assumptions of a normal distribution of residuals, equal variance, and linearity. All assumptions were approximately met (see Appendix L). Running the linear model, we did not see any significant influence of sound on the perceived strength of the emotion. Also, age did not have a significant influence on the perceived strength of the emotions (see Table 5.4). The mean scores of the strength of the perceived emotions were for no sounds 3.17 (SD = 1.93), for beeping sounds 3.14 (SD = 1.77), and for musical sounds 3.19 (SD = 1.69).

Table 5.4: Linear Mixed Model for Perceived Strength of Emotion.

Coefficients	B	Std. Error	t	Sig.
Intercept	2.78	0.27	10.40	.000
No sound	0.01	0.18	0.05	.957
Beeping	-0.03	0.18	-0.17	.864
Age	0.01	0.01	1.73	.084

Note. The musical utterance condition was the reference category.

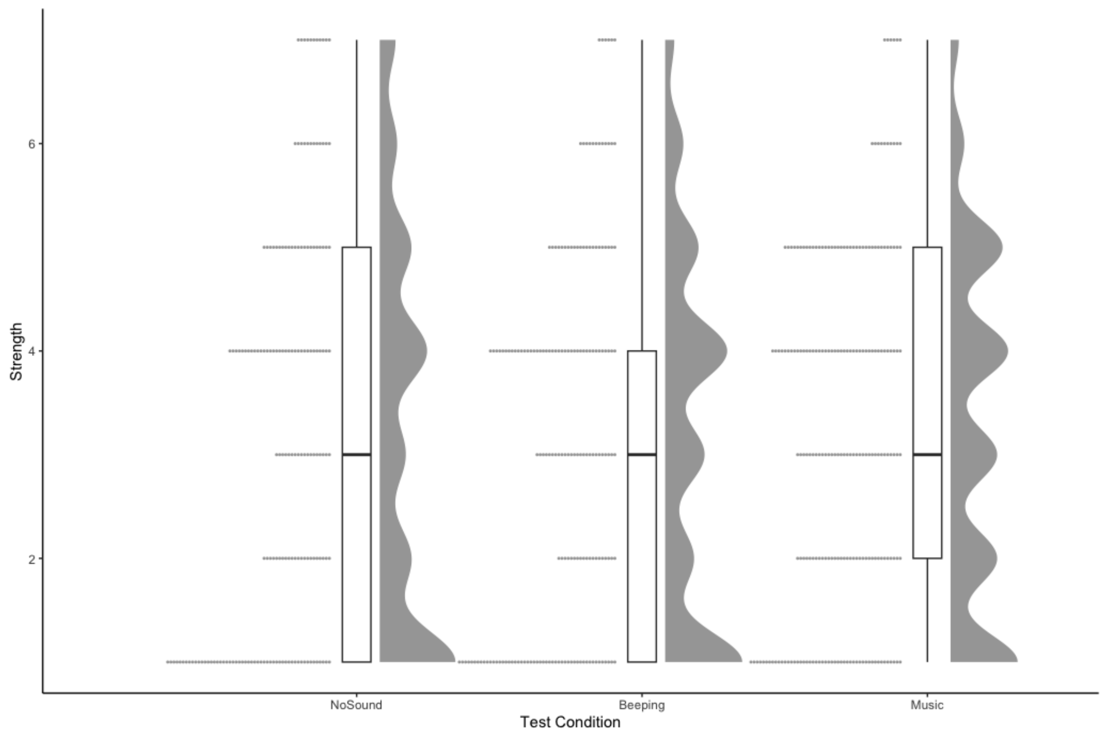


Figure 5.4: Raincloud plot showing the perceived strength of the emotion for every condition.

5.3 Empathy

5.3.1 Cognitive Empathy

In Figure 5.5, the distribution of the experienced levels of cognitive empathy per condition can be seen. The influence of sound on cognitive empathy was tested by running a linear mixed model. Here as well, we included age, gender, and experience with robots as covariates separately. None of the covariates had a significant influence on the level of cognitive empathy. However, including age in the model did improve the model fit ($\chi 2(1)$ = 7.43, $p \leq .001$, see Appendix M), and hence it was included in the model. We checked the assumptions of normal distribution of residuals, equal variance, and linearity. The former one was not met. However, the latter two were met (see Appendix O) and inferences from the models are generally robust against these violations. Hence the model still seemed suitable. The linear mixed model showed that there is a significant difference between the no-sound condition and the musical sound condition regarding the level of cognitive empathy. People in the musical sound condition experience a significantly higher level of cognitive empathy. However, there was no significant difference between the beeping sound and musical utterance condition regarding the level of cognitive empathy. Additionally, the effect of age was also not significant (see Table 5.5). The mean scores for cognitive empathy were 3.11 (SD = 1.62) for the no-sound condition, 3.38 (SD = 1.59) for the beeping sound condition, and 3.47 (SD = 1.80) for the musical sound condition.

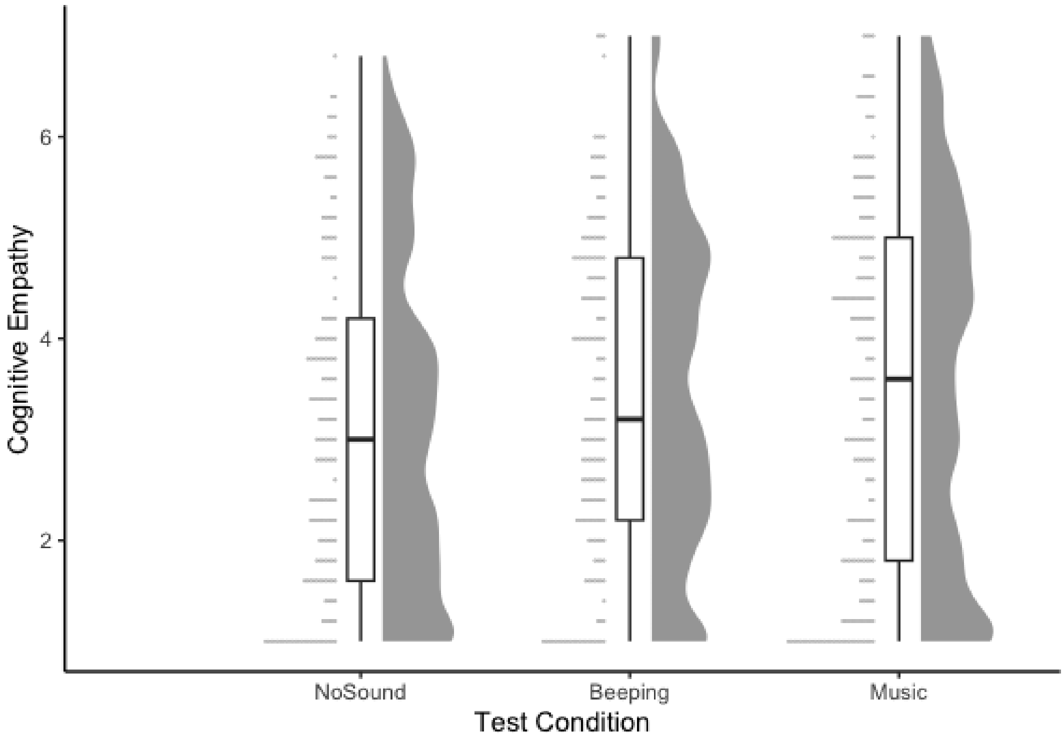


Figure 5.5: Raincloud plot showing the distribution of the level of cognitive empathy experienced per condition.

Table 5.5: Linear mixed model for cognitive empathy.

Coefficients	B	Std. Error	t	Sig.
Intercept	3.69	0.25	14.73	.000
No sound	-0.38	0.17	-2.18	.030
Beeping	-0.10	0.17	-0.58	.562
Age	-0.01	0.01	-1.01	.314

Note. The musical utterance condition was the reference category.

5.3.2 Affective Empathy

Figure 5.6 shows the levels of affective empathy perceived by the participants per sound condition. We ran a linear mixed model on the levels of affective empathy with the sound conditions as a fixed predictor variable and scenarios as a random effect. Again, we tested the influence of age, gender, and experience with robots as separate covariates. None of them had a significant influence on affective empathy. However, age significantly improved the model ($\chi 2$(1) = 7.63, *p* = .005, see Appendix M) and hence, was included in the model. The assumptions of normal distribution of residuals, equal variance, and linearity were found to be approximately met (see Appendix N) and therefore a linear mixed model fitted the data. We saw that the sound condition did not have a significant effect on the level of affective empathy experienced. Also, age did not have a significant influence on the level of affective empathy (see Table 5.6). The mean scores for affective empathy were 2.56 (SD = 1.34) for the no sound condition, 2.50 (SD = 1.21) for the beeping sound condition, and 2.68 (SD = 1.35) for the musical sound condition.

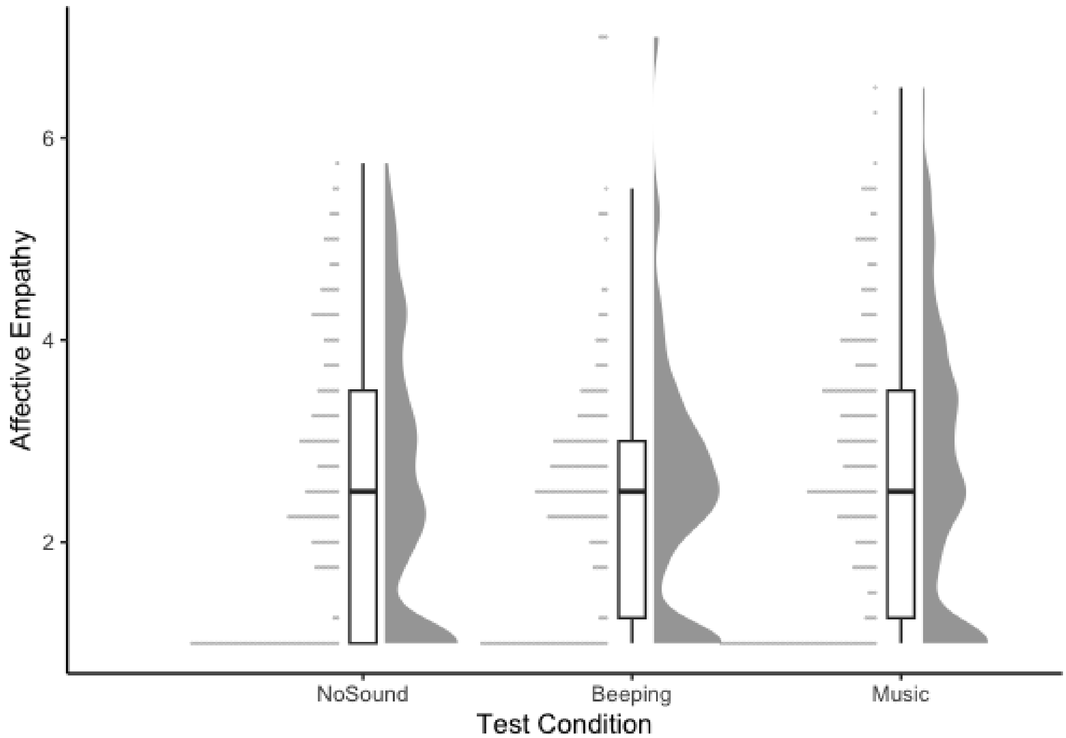


Figure 5.6: Raincloud plot showing the levels of affective empathy felt by the participants in every sound condition.

Table 5.6: Linear Mixed Model for Affective Empathy.

Coefficients	B	Std. Error	t	Sig.
Intercept	2.88	0.19	14.80	.000
No sound	-0.14	0.13	-1.04	.298
Beeping	-0.20	0.13	-1.48	.140
Age	-0.01	0.00	-1.15	.250

Note. The musical utterance condition was the reference category.

5.4 Prosocial Behavior

Figure 5.7 shows the distribution of intended prosocial behavior for every sound condition. We ran an ordinal logistic mixed model to test the effect of sound on prosocial behavior including the scenarios as random effects. As before, we tested whether the covariates age, gender, and experience with robots should be included in the model. Age and experience with robots did not have a significant influence on the level of prosocial behavior. Being male significantly increased the log odd ratio of being in a higher category compared to female participants. However, including gender in the model did not improve the model fit ($\chi 2(1)$ = 5.01, *p* =.081) and so did age ($\chi 2(1)$ = 0.09, *p* =.769) and experience with robots also not ($\chi 2(1)$ = 1.38, *p* =.241, see Appendix P). Therefore, we did not include any covariate in the model. From the analysis, we saw that sound does not have a significant influence on the level of intended prosocial behavior (see Table 5.7).

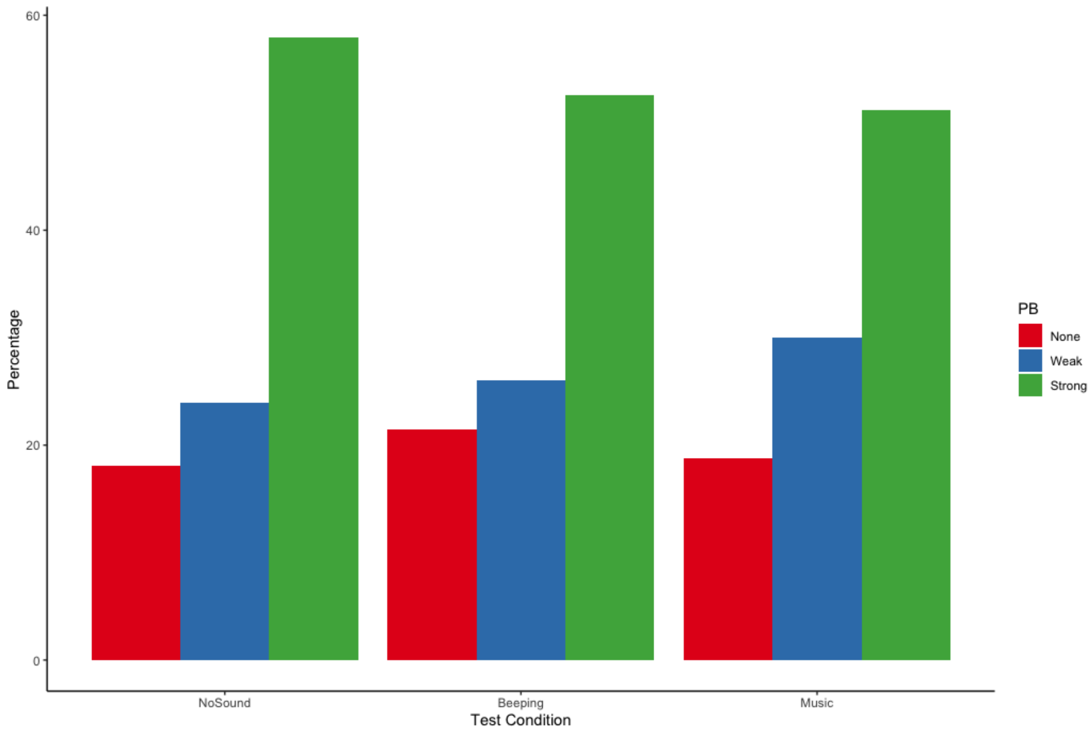


Figure 5.7: Bar chart showing the distribution of prosocial behavior for every condition.

Table 5.7: Ordinal Logistic Mixed Model for Prosocial Behavior.

Coefficients	B	Std. Error	t	Sig.
No sound	0.21	0.20	1.08	.282
Beeping	-0.00	0.19	-0.01	.991

Note. The musical utterance condition was the reference category.

Chapter 6

Discussion

6.1 Discussion

In the current study, we looked at the influence of sounds on legibility, perceived emotions, empathy, and prosocial behavior in response to the communication of a robot. Some of our results hint at the role different kinds of sounds play. When using beeping sounds, people are more confident in the interpretation of the communication compared to musical utterances. This possibly indicates, that more complex sounds can lead to more confusion with regard to their interpretation. Next to that, people are more likely to perceive emotions evoking empathy in the musical utterance sound condition compared to the beeping sound condition. The insight that music conveys emotions hence seems to translate also to HRI. The deliberate design of rather negative emotions seemed to be successful. Lastly, people in the musical utterance condition experienced significantly more cognitive empathy compared to people in the no sound condition. People seem to be more consciously processing the state the robot communicates to be in.

However, other results also indicated no differences between the conditions. We found no difference in the legibility of the robot's communication depending on the used sound. Overall, a high number of participants understood the robot's behavior and intentions. Hence, there was not much room for musical utterances to even improve this further. Additionally, our results suggested that there is no difference in confidence between no sounds and musical utterances. Accordingly, adding more complex sounds seems to diminish the positive effect of simpler sounds again. Also, our data do not indicate significant differences between the no sound and the musical utterance condition regarding the perception of emotions. This comes rather surprising, especially seeing the significant difference between musical utterances and beeping. Regarding empathy, the sound did not influence the levels of affective empathy people experienced in response to the robot. All sounds evoked a similar experience of empathy which was rather low. Similarly, we did not observe significant differences between the beeping condition and the musical utterance condition with regard to cognitive empathy. Musical utterances do not seem to offer additional information for the understanding and interpretation of the emotions communicated by the robot. Finally, we could not find a difference in the way people engage in prosocial behavior depending on the sound condition. Interestingly, the levels of prosocial behavior were high in all conditions. Hence the question arises whether they would also be similarly high in real-life situations. In the following sections, we will have a closer look at the meaning of our results and how they relate to already existing research.

6.1.1 Legibility

Based on the results, we cannot accept the first hypothesis, that when using musical utterances, the communication of the robot is more legible compared to beeping sounds and no sounds. In all conditions, the communication of the robot was about equally legible and most people (around 75%) were able to correctly interpret what the robot was trying to communicate (see Figures 5.1). This is most likely a result of the fact that the situations in the videos were not ambiguous. We created the videos to display the problem of the robot. Our goal was not to enable different kinds of interpretations of the videos. Instead, by measuring legibility, we wanted to ensure that the videos and especially the communication of the robot were understood by the participants. This was the case for the great majority which is important when looking further at the other variables. If the majority of the participants would have interpreted the situation and communication of the robot completely differently, this possibly would have had an influence on the other variables as well. Especially for resolving the problem of the robot, the right interpretation of the communication and the problem in the first place is important.

However, we saw a significant difference between beeping and musical sounds regarding confidence in the interpretation of the communication. Unexpectedly, people in the beeping sound condition were significantly more confident than people in the musical utterance condition. Hence, the beeping can be considered more legible. The beeping sounds are non-linguistic utterances. Accordingly, they communicate some level of contextual information. Non-linguistic utterances are generally said to be easily understandable [26]. Often, they are designed very simplistic and hence make comprehension by humans fast and easy. Therefore, they are very suitable for, among other things, informational alerts [26]. Our results indicate simple sounds like non-linguistic utterances can already be sufficient valuable additional information to convey a message. Especially as the scenarios in the current study are already well understood by the majority of the participants, simple sounds further underline these and can be regarded as additional informational channels that can be comprehended without much effort needed from the humans.

Nonetheless, the confidence in the interpretation did not differ significantly between the no sound and the musical utterance condition. The levels of confidence were about the same. Hence, we could not find a difference in legibility here. One reason for this could be that the participants were surprised to hear music as means of communication. As described before, the use of musical utterances beyond the communication of emotions is to date rather rare. Hence they are not familiar with the use of music for the communication of information. Huron [66] argued that when we are more familiar with certain music we are better able to predict events and hence have an increased understanding of its meaning. In turn, this also means that if music occurs in a more unfamiliar setting, the additional information cannot be easily interpreted by the listener. Additionally, there is a large body of evidence for music to convey basic emotions (for example [67] [68] [69]). However, beyond the basic emotions, the expression of other information through musical utterances becomes more complex, and different people seem to have different interpretations of the music. More complex information conveyed by music requires more intrinsic features of the music. A greater combination and variation of musical properties have to be used and hence longer sequences of music are needed to effectively communicate this distinguishable to other meanings [68]. The musical sequences in the current research were relatively short. Additionally, they were meant to convey relatively complex information. Therefore, they might have introduced an additional channel of communication for the robot. However, musical utterances introduced additional complexity, leaving room for different interpretations. Accordingly, they possibly require more cognitive efforts to be processed compared to simpler beeping sounds. Hence, musical utterance diminished the benefits sounds can have as means of communication. This can be also confirmed by the responses to the open questions. Some participants interpreted the musical utterances for example as greetings. These interpretations did not occur in the beeping and no sound condition. Accordingly, this indicates that musical utterances similar to other complex sounds can be interpreted in many ways and introduce ambiguity.

With regard to the confidence people had in their interpretations of the situation, it is important to note that we did not account for wrong interpretations of the communications. This means that there is a chance that people understood the communication of the robot wrong but were still very confident in their understanding. However, this could even be regarded as less legible robot communication than having low confidence in the correct interpretation. Nonetheless, in the current study, this issue possibly does not have a big influence. In the first place, a similar amount of people in all conditions interpreted the communication wrong and there was no big difference there. Hence, the case that people gave a strong rating for a wrong interpretation most likely occurred similarly often in all conditions, reducing the influence it has on our results regarding the differences in confidence. Still, in future studies, this should be considered more carefully by accounting for high but wrong confidence levels like for example in [70] and [71]. Also, this influence should be kept in mind for the current interpretation of the results.

6.1.2 Perceived Emotions

The results from our study partly confirm the second hypothesis regarding the increased perception of emotions evoking empathy when communicating through musical utterances compared to beeping and no sounds. Participants were more likely to perceive empathy-evoking emotions in the musical utterance condition compared to the beeping sound condition. Hence, participants are also more likely to perceive more emotions generally in the musical utterance condition compared to the beeping condition. For the comparison between the no sound condition and the musical utterance condition, our results were not significant. While not significant, our results hint in the direction that people were more likely to perceive emotions in a lower category in the no sound conditions, hence, emotions that are less related to empathy. Though we cannot draw concrete conclusions and this has to be cautiously interpreted, this is in line with the findings of [35]. They demonstrated that music is an important means for the communication of emotions within robots. People were strongly influenced by the emotions conveyed through music. Similarly, in another study, the hypothesis to use musical sounds together with other modalities was confirmed. Participants were able to perceive emotions in response to a robot using music, among others, to convey the emotions [36]. Therefore, our findings, that overall more emotions are perceived in the musical utterance condition, is in line with already existing research. Additionally, the fact that the emotions perceived are beneficial for empathy is a result of the sound design. The designers were instructed to create sounds with the goal of evoking empathy. They mostly had negative emotions, such as distress and sadness, in their mind. These are negative emotions that are best in evoking situational empathy [49].

However, it is important to note that, generally, the level of perceived emotions was rather low, indicating that the situation was not very emotional, and sounds did not add much to this. In all conditions, more than half of the participants rated the emotions of the robot as neutral and in the no sound and beeping sound conditions it was even more than 60% of the participants. Looking at the answers to the open questions regarding perceived emotions, a great number of participants mentioned that robots do not have any feelings or emotions and hence there are no emotions to perceive. This a very conscious cognitive process, which was triggered by asking for their perception. However, generally, the perception of emotions happens more automatically and unconsciously [72]. Therefore, the perception of emotions is possibly different in real life. Additionally, the musical sounds were only able to counteract this to a limited extent. This might be a result of the freedom the sound designers had in the design of the sounds. Although the sound designers had clear instructions on what the message to convey is, we did not give any instructions on how to do it. They had to split their focus and creativity to convey the intention while also being emotional. However, specific features of music directly map to perceiving specific emotions in music [34]. Yet, to convey the intentions as well these might have not been made use of fully. Certain intentions might have been conveyed through different musical features than the emotions the sound designers had in mind. Therefore, by making the trade-off between communicating intentions and emotions, the intentions possibly have interfered with the communication of emotions.

6.1.3 Empathy

We can only partly confirm hypothesis 3.1, which states that people experience more cognitive empathy when hearing musical utterances compared to beeping sounds and no sounds. We observed a significant difference between musical utterances and no sound but no significant difference between beeping sounds and musical utterances. Additionally, with regard to hypothesis 3.2, the results did not confirm that there is a significant difference in the levels of affective empathy between the beeping sound condition and the musical utterance condition. Overall, in order to feel empathy more generally, cognitive and affective empathy both have to be present and interact [43]. Hence, according to our results, the overall levels of empathy are similar in all conditions, because at least the levels of affective empathy do not differ significantly between the conditions.

However, looking at the results of the perceived emotions, the results with regard to affective and cognitive empathy do not come as a surprise. As described earlier, the majority of participants in each condition did not perceive any emotions in the communication of the robot and described it as neutral. This was a result of the cognitive process behind describing emotions and the experience in real life most likely is different. Nonetheless, empathy is a response to the perception of emotions within others. The observer perceives an emotion displayed by a target and based on this might experience empathy. Although the emotions of the observer do not need to be the same as the emotions experienced by the target, in order to elicit empathy an emotional stimulus is required [43]. Also, as described in 6.1.2, part of the participants seemed to be influenced by their knowledge that robots do not have any emotions. Possible cognitive empathetic processes influenced affective empathetic responses as well. Generally, the more conscious cognitive empathy can have an impact on affective empathy [43]. Therefore, the conscious knowledge that the robots do not have any emotions or feelings might have had an impact on the experienced affective empathy. However, with the data of the current study, we cannot further look into this. So overall, in the current study, the majority of participants did not perceive any emotions in the communication of the robot and hence no emotional stimulus was given to them. This results in low levels of empathy and similar levels of empathy across conditions.

Even in the case that there are differences in the perceived emotions, there are no significant differences in their perceived strength. In previous studies, a strong relationship between the intensity of the emotion display and the level of empathy felt was observed. If emotions are expressed more intensely by the target, the observer is likely to experience higher levels of empathy [9]. However, in the current study, all emotions are perceived at a similar level of intensity and there is no significant difference in the perceived strength. Hence, this also does not enhance empathy in one condition compared to another one.

Eventually, familiarity also has an influence on empathy. If we are unfamiliar with the other person or object we are likely to experience lower levels of empathy [57]. The participants indicated that they have rather low levels of experience with robots and hence most likely are not familiar with them. Especially in hospitals, transportation robots are not yet widely introduced. Therefore, the whole scenario seems to be quite unfamiliar to the participants and hence, reduces the likelihood of empathy being experienced.

Interestingly, our results showed a significant difference in the levels of cognitive empathy between the no sound and the musical utterance condition. Cognitive empathy is the more controlled process with regard to empathy and can be consciously manipulated by people. Part of cognitive empathy is for example perspective taking [43]. Hence, with musical utterances, people might be better able to put themselves into the shoes of the robot. However, our current study does not allow us to look at whether this is true and what the possible reasons are. Therefore this should be further looked at in the future.

6.1.4 Prosocial Behavior

The current study does not confirm the fourth hypothesis that there is a significant difference in the level of prosocial behavior between the use of musical sounds and the use of beeping or no sounds by a robot. Again, looking at the results with regard to empathy, these results concerning prosocial behavior are not surprising. People tend to engage in more prosocial behavior when they feel more empathetic. Kamas and Preston [59] demonstrated that empathy and prosocial behavior are highly positively correlated. Similarly, in another study, it was demonstrated that when people feel more empathetic toward a robot, they are more likely to help that robot when it is stuck [21]. As participants in the current study do not experience significant differences in the levels of empathy in the different conditions, we also cannot expect differences in the levels of prosocial behavior they intend to engage in.

Additionally, it can be seen that overall, the level of prosocial behavior is very high across all conditions. A reason for this might be socially desirable answers. These are answers that are framed as good answers through social norms [73]. Especially when talking about prosocial behavior, increasing one's own reputation and status plays an important role. Hardy and Van Vugt [74] showed in their study, that people who behave more altruistically, experience more social benefits and are associated with a higher social status. Hence, people possibly have been aware that answering with some prosocial behavior is more socially desirable and beneficial for them. Although in the instructions we mentioned that there are no right or wrong answers several times, the idea of showing socially desirable behaviors might have still influenced the participants, at least subconsciously. Accordingly, as we merely measure the intention to engage in prosocial behavior rather than prosocial behavior itself, the reality might be different. When people are actually in the situation, their actual behavior is likely to be different from what they imagine they would be doing [75]. Therefore, the overall level of intended prosocial behavior was high in our study. However, this might have been due to social desirability, and in the actual situation, the level of prosocial behavior possibly differs due to the intention-behavior discrepancy.

6.1.5 Age and Experience with Robots

Age and experience with robots had significant influences on some of the variables in contrast to gender. People that were more experienced with robots and that were older, had a higher level of confidence in the answer to the open question about the communication of the robot. Additionally, those people also perceive more emotions in the communication of the robot. Especially for experience, this can be expected. People that have more experience with robots, understand them better. Horstmann and Krämer [76] found a relationship between previous experiences with robots and the expectations they have. People assess the abilities and skills of robots based on the experience they have with robots. This enhances a realistic understanding of this technology [76]. Therefore, the interpretation of the behavior seems to be easier for them. They have a better understanding of robots and can integrate the behavior into the bigger picture.

However, for age, this result is surprising. One would expect that people who are older are more uncertain about technology and therefore, have more difficulty understanding robots. Generally, older people experience greater difficulty using unfamiliar technologies and have the feeling of lower self-efficacy [77]. Our results show the opposite. Yet, the differences are very small. Age was not a variable of interest in our study and hence we did not purposefully design for it. However, in the future, it might be interesting to do this and check what role age plays with regard to legibility and emotions. A similar note holds for experience with robots and it seems interesting to further investigate this.

6.2 Limitations

Importantly, the current study only concerned situations in which the cause of the problem was external to the robot. Initially, we also tried to include an internal problem of the robot (losing navigation) to take those situations also into account. However, during the sound design, as well as during the pilot study this scenario turned out to be more complicated than the other scenarios. When the cause of the problem is internal, the sound designers had no reference point for their sounds. This made the sound design difficult. Additionally, communicating internal problems without the use of words or any other communication channels than sound is very difficult. Hence, the current results are only applicable to situations in which the cause of the problem is external to the robot.

Additionally, there might have been a mismatch between the sounds and the embodiment of the robot. The sounds have been designed for the robot that is developed and designed in the Harmony project (see Figure 3.1). However, due to scheduling issues, this one was not available for the study. Therefore, another robot was used (see Figure 4.2). Although we tried our best to select a very similar robot, this can still have resulted in a mismatch between the robot's appearance and the sounds designed. However, matching the appearance and voice or sound of a robot is highly important in HRI. A mismatch between the two can results in user expectations not being met and therefore have adverse effects on the interaction and evaluation [78]. Even though we aimed for a high similarity, we cannot rule out that there was a mismatch between the sounds and the appearance of the robot and hence that this mismatch had an effect on our results.

Next to that, our results should be interpreted with caution. Due to time and resource limitations, we conducted an online study, where the participants saw videos of an acted situation. Therefore, the participants did not experience the situation in real life and had merely an observer position. This possibly has an influence on the levels of legibility, perceived emotions, and experienced empathy. Also, the study measures the intention to engage in prosocial behavior compared to actual prosocial behavior itself. However, there can be a difference between what people intend to do and what they are actually doing. Therefore, we can only draw conclusions about empathy and prosocial behavior to a limited extent. Rather the results should be regarded as an indication but real-life studies are needed to draw concrete conclusions.

Also, it is important to acknowledge that we did not take other variables into account regarding empathy and prosocial behavior that in previous research have been proven to have an influence on them. For example, familiarity influences the level of experienced empathy. If people are more familiar with a robot, they are likely to be more empathetic towards it [57]. Similarly, when people experience higher levels of distress in a situation they are more likely to engage in prosocial behavior [79]. Accounting for these and other variables would have been out of the scope of the current study. However, we cannot rule out that they are important to consider in the sound design and that they still had an influence on our results that we did not see.

Finally, we presented the videos of the scenarios in the same order to all participants. Unfortunately, SurveyMonkey did not allow to randomize the order of the videos. This might have introduced an order effect. Especially with visual stimuli, primacy effects can be expected in the studies. Visual stimuli that are shown at the beginning might be used as a standard for comparison and are differently, more intensively processed by people [80]. Therefore, when seeing the latter visual stimuli, they are possibly framed by what has already been perceived beforehand. Accordingly, we cannot rule out that responses for the latter videos might have been biased by the first videos. As the videos were always in the same order, always the responses to the same videos might be biased.

Chapter 7

Conclusions and Recommendations

In our study, we tried to find an answer to the question *'To what extent are musical utterances able to communicate intention and emotions that result in empathy and prosocial behavior toward a robot?'*. Musical utterances were able to convey more empathy-evoking emotions compared to beeping sounds. The results also hinted at more emotions perceived in musical utterances compared to no sounds, but our results were not significant. With regard to empathy, the musical utterances only made a significant difference for cognitive empathy when compared to the no sound condition.

We made use of scenarios that were not ambiguous and easy to interpret. Hence, musical utterances did not add much to the legibility. Moreover, the musical utterances might have introduced an additional channel of information that is not easy to interpret when compared to beeping and might add confusion. We did not observe an influence on the levels of cognitive empathy experienced between beeping sounds and musical sounds. Also, music did not make a significant difference in the experience of affective. Accordingly, given the mediating role of empathy, we could also not expect a difference in the level of prosocial behavior which we also did not observe.

From these results, it can be concluded that in unambiguous situations and scenarios in the current use case, no sound is required to communicate the intentions of the robot and the situation already entails enough information on its own. Beeping sounds can be used to increase the confidence of the users in the understanding of the robot. Musical utterances are mainly useful in the communication of emotions. However, with our current study, we cannot draw concrete conclusions about whether this leads also to increased empathy and willingness to engage in prosocial behaviors. A main reason for this is the nature of the current study, which gives good reasons for future research to further look into the effects musical utterances can have, especially with regard to empathy and prosocial behavior.

In future research, the sounds for the robot should be designed more considerately with the context in mind and based on already existing theory. This would ensure a closer fit of the sounds. Additionally, a real-life study should be conducted to get more insights into the effects of musical utterances on empathy and prosocial behavior. With our online study, we could only investigate these variables to a limited extent. Finally, when looking at the use of musical sounds to evoke empathy and prosocial behavior, other variables that we for the current study disregarded but that play an important role should be taken into account to get a more holistic picture.

Bibliography

[1] A. G. Ozkil, Z. Fan, S. Dawids, H. Aanes, J. K. Kristensen, and K. H. Christensen, “Service robots for hospitals: A case study of transportation tasks in a hospital,” in *2009 IEEE international conference on automation and logistics*. IEEE, 2009, pp. 289–294.

[2] J. X. Liu, Y. Goryakin, A. Maeda, T. Bruckner, and R. Scheffler, “Global health workforce labor market projections for 2030,” *Human resources for health*, vol. 15, no. 1, pp. 1–12, 2017.

[3] S. Bandyopadhyay, R. E. Baticulon, M. Kadhum, M. Alser, D. K. Ojuka, Y. Badereddin, A. Kamath, S. A. Parepalli, G. Brown, S. Iharchane *et al.*, “Infection and mortality of healthcare workers worldwide from covid-19: a systematic review,” *BMJ global health*, vol. 5, no. 12, p. e003097, 2020.

[4] Harmony. (2021) About us. [Online]. Available: https://harmony-eu.org/project/about/

[5] A. Paiva, I. Leite, H. Boukricha, and I. Wachsmuth, “Empathy in virtual agents and robots: A survey,” *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 3, pp. 1–40, 2017.

[6] A. Macaskill, J. Maltby, and L. Day, “Forgiveness of self and others and emotional empathy,” *The Journal of social psychology*, vol. 142, no. 5, pp. 663–665, 2002.

[7] J. Feng, J. Lazar, and J. Preece, “Empathy and online interpersonal trust: A fragile relationship,” *Behaviour & Information Technology*, vol. 23, no. 2, pp. 97–106, 2004.

[8] L. R. Welp and C. M. Brown, “Self-compassion, empathy, and helping intentions,” *The Journal of Positive Psychology*, vol. 9, no. 1, pp. 54–65, 2014.

[9] M. L. Hoffman, *Empathy and moral development: Implications for caring and justice*. Cambridge University Press, 2001.

[10] M. W. Kraus, "Voice-only communication enhances empathic accuracy." *American Psychologist*, vol. 72, no. 7, pp. 644 – 654, 2017.

[11] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech communication*, vol. 40, no. 1-2, pp. 227–256, 2003.

[12] S. Yilmazyildiz, R. Read, T. Belpeame, and W. Verhelst, "Review of semantic-free utterances in social human–robot interaction," *International Journal of Human-Computer Interaction*, vol. 32, no. 1, pp. 63–85, 2016.

[13] C. Zaga, R. A. De Vries, S. J. Spenkelink, K. P. Truong, and V. Evers, "Help-giving robot behaviors in child-robot games: Exploring semantic free utterances," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 541–542.

[14] R. Read and T. Belpaeme, "How to use non-linguistic utterances to convey emotion in child-robot interaction," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 219–220.

[15] K. Seaborn and P. Pennefather, "Neither "hear" nor "their": Interrogating gender neutrality in robots," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 1030–1034.

[16] F. A. Robinson, M. Velonaki, and O. Bown, "Crafting the language of robotic agents: A vision for electroacoustic music in human–robot interaction," *Organised Sound*, pp. 206–218, 2022.

[17] E.-S. Jee, Y.-J. Jeong, C. H. Kim, and H. Kobayashi, "Sound design for emotion and intention expression of socially interactive robots," *Intelligent Service Robotics*, vol. 3, pp. 199–206, 2010.

[18] R. Savery, R. Rose, and G. Weinberg, "Finding shimi's voice: fostering human-robot communication with music and a nvidia jetson tx2," in *Proceedings of the 17th Linux Audio Conference*, vol. 5, 2019, pp. 101–105.

[19] A. M. Rosenthal-von der Pütten, N. C. Krämer, L. Hoffmann, S. Sobieraj, and S. C. Eimler, "An experimental study on emotional reactions towards a robot," *International Journal of Social Robotics*, vol. 5, no. 1, pp. 17–34, 2013.

[20] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll, "Learning by feeling: Evoking empathy with synthetic characters," *Applied Artificial Intelligence*, vol. 19, no. 3-4, pp. 235–266, 2005.

[21] D. G. Holm, R. P. Junge, M. Østergaard, L. Bodenhagen, and O. Palinko, "What will it take to help a stuck robot? exploring signaling methods for a mobile robot," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 797–801.

[22] S. Yilmazyildiz, D. Henderickx, B. Vanderborght, W. Verhelst, E. Soetens, and D. Lefeber, "Emogib: emotional gibberish speech database for affective human-robot interaction," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 163–172.

[23] R. Read and T. Belpaeme, "People interpret robotic non-linguistic utterances categorically," *International Journal of Social Robotics*, vol. 8, pp. 31–50, 2016.

[24] ——, "Situational context directs how people affectively interpret robotic non-linguistic utterances," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 2014, pp. 41–48.

[25] H. Wolfe, M. Peljhan, and Y. Visell, "Singing robots: How embodiment affects emotional responses to non-linguistic utterances," *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 284–295, 2017.

[26] J. S. Lee, M. F. B. Abbas, C. K. Seow, Q. Cao, K. P. Yar, S. L. Keoh, and I. McLoughlin, "Non-verbal auditory aspects of human-service robot interaction," in *2021 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, 2021, pp. 1–5.

[27] C. Breazeal, *Designing sociable robots*. MIT press, 2002.

[28] K. R. Scherer, "Expression of emotion in voice and music," *Journal of voice*, vol. 9, no. 3, pp. 235–248, 1995.

[29] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human-Computer Interaction*, vol. 19, no. 1, pp. 167–177, 1986.

[30] J. Sterkenburg, M. Jeon, and C. Plummer, "Auditory emoticons: Iterative design and acoustic characteristics of emotional auditory icons and earcons," in *Advanced Interaction Modalities and Techniques: 16th International Conference*. Springer, 2014, pp. 633–640.

[31] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human–Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[32] P. Larsson, A. Opperud, K. Fredriksson, and D. Västfjäll, "Emotional and behavioural response to auditory icons and earcons in driver-vehicle interfaces," in *Proc. 21st International Technical Conference on Enhanced Safety of Vehicles, Germany*, 2009.

[33] K. Hevner, "The affective character of the major and minor modes in music," *The American Journal of Psychology*, vol. 47, no. 1, pp. 103–118, 1935.

[34] P. N. Juslin and E. Lindström, "Musical expression of emotions: Modelling listeners' judgements of composed and performed features," *Music Analysis*, vol. 29, no. 1-3, pp. 334–364, 2010.

[35] E.-S. Jee, C. H. Kim, S.-Y. Park, and K.-W. Lee, "Composition of musical sound expressing an emotion of robot based on musical factors," in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2007, pp. 637–641.

[36] R. Savery, R. Rose, and G. Weinberg, "Establishing human-robot trust through music-driven robotic emotion prosody and gesture," in *2019 28th IEEE international conference on robot and human interactive communication (RO-MAN)*. IEEE, 2019, pp. 1–7.

[37] J. G. Bloodsworth, "Legibility of print." 1993.

[38] C. Lichtenthäler and A. Kirsch, "Legibility of robot behavior: A literature review," 2016.

[39] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 301–308.

[40] G. Angelopoulos, A. Rossi, C. Di Napoli, and S. Rossi, "You are in my way: Non-verbal social cues for legible robot navigation behaviors," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 657–662.

[41] C. Lichtenthäler and A. Kirsch, "Towards legible robot navigation-how to increase the intend expressiveness of robot navigation behavior," in *International Conference on Social Robotics-Workshop Embodied Communication of Goals and Intentions*, 2013.

[42] L. Wispé, "The distinction between sympathy and empathy: To call forth a concept, a word is needed." *Journal of personality and social psychology*, vol. 50, no. 2, p. 314, 1986.

[43] B. M. Cuff, S. J. Brown, L. Taylor, and D. J. Howat, "Empathy: A review of the concept," *Emotion review*, vol. 8, no. 2, pp. 144–153, 2016.

[44] R. P. Ang and D. H. Goh, "Cyberbullying among adolescents: The role of affective and cognitive empathy, and gender," *Child Psychiatry & Human Development*, vol. 41, no. 4, pp. 387–397, 2010.

[45] A. Main, E. A. Walle, C. Kho, and J. Halpern, "The interpersonal functions of empathy: A relational perspective," *Emotion Review*, vol. 9, no. 4, pp. 358–366, 2017.

[46] T. Norfolk, K. Birdi, and D. Walsh, "The role of empathy in establishing rapport in the consultation: a new model," *Medical education*, vol. 41, no. 7, pp. 690–697, 2007.

[47] J. D. Wondra and P. C. Ellsworth, "An appraisal theory of empathy and other vicarious emotional experiences." *Psychological review*, vol. 122, no. 3, pp. 411–428, 2015.

[48] K. Zhou, L. M. Aiello, S. Scepanovic, D. Quercia, and S. Konrath, "The language of situational empathy," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 1–19, 2021.

[49] J. Jauniaux, M.-H. Tessier, S. Regueiro, F. Chouchou, A. Fortin-Côté, and P. L. Jackson, "Emotion regulation of others' positive and negative emotions is related to distinct patterns of heart rate variability and situational empathy," *PloS one*, vol. 15, no. 12, p. e0244427, 2020.

[50] J. Zaki, N. Bolger, and K. Ochsner, "Unpacking the informational bases of empathic accuracy." *Emotion*, vol. 9, no. 4, pp. 478–487, 2009.

[51] A. De Waele, L. Schoofs, and A.-S. Claeys, "The power of empathy: the dual impacts of an emotional voice in organizational crisis communication," *Journal of Applied Communication Research*, vol. 48, no. 3, pp. 350–371, 2020.

[52] M. Fabri, D. Moore, and D. Hobbs, "Expressive agents: Non-verbal communication in collaborative virtual environments," *Proceedings of Autonomous Agents and Multi-Agent Systems (Embodied Conversational Agents)*, 2002.

[53] C. Regenbogen, D. A. Schneider, A. Finkelmeyer, N. Kohn, B. Derntl, T. Kellermann, R. E. Gur, F. Schneider, and U. Habel, "The differential contribution of facial expressions, prosody, and speech content to empathy," *Cognition & emotion*, vol. 26, no. 6, pp. 995–1014, 2012.

[54] T. S. Polzin and A. Waibel, “Detecting emotions in speech,” in *Proceedings of the CMC*, vol. 16. Citeseer, 1998.

[55] S. Brave, C. Nass, and K. Hutchinson, “Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent,” *International journal of human-computer studies*, vol. 62, no. 2, pp. 161–178, 2005.

[56] H. Abdollahi, M. Mahoor, R. Zandie, J. Sewierski, and S. Qualls, “Artificial emotional intelligence in socially assistive robots for older adults: A pilot study,” *IEEE Transactions on Affective Computing*, 2022.

[57] J. K. Malinowska, “Can i feel your pain? the biological and socio-cognitive factors shaping people’s empathy with social robots,” *International Journal of Social Robotics*, vol. 14, no. 2, pp. 341–355, 2022.

[58] R. Oliveira, P. Arriaga, F. P. Santos, S. Mascarenhas, and A. Paiva, “Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour,” *Computers in Human Behavior*, vol. 114, p. 106547, 2021.

[59] L. Kamas and A. Preston, “Empathy, gender, and prosocial behavior,” *Journal of Behavioral and Experimental Economics*, vol. 92, p. 101654, 2021.

[60] N. Eisenberg and P. A. Miller, “The relation of empathy to prosocial and related behaviors.” *Psychological bulletin*, vol. 101, no. 1, pp. 91–119, 1987.

[61] A. C. Rumble, P. A. Van Lange, and C. D. Parks, “The benefits of empathy: When empathy may sustain cooperation in social dilemmas,” *European Journal of Social Psychology*, vol. 40, no. 5, pp. 856–866, 2010.

[62] F. B. Parmentier, G. Elford, C. Escera, P. Andrés, and I. San Miguel, “The cognitive locus of distraction by acoustic novelty in the cross-modal oddball task,” *Cognition*, vol. 106, no. 1, pp. 408–432, 2008.

[63] N. Ahmadpour, S. Pedell, A. Mayasari, and J. Beh, “Co-creating and assessing future wellbeing technology using design fiction,” *She Ji: The Journal of Design, Economics, and Innovation*, vol. 5, no. 3, pp. 209–230, 2019.

[64] L. H. Kim and S. Follmer, “Generating legible and glanceable swarm robot motion through trajectory, collective behavior, and pre-attentive processing features,” *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 3, pp. 1–25, 2021.

[65] R. L. Reniers, R. Corcoran, R. Drake, N. M. Shryane, and B. A. Völlm, "The qcae: A questionnaire of cognitive and affective empathy," *Journal of personality assessment*, vol. 93, no. 1, pp. 84–95, 2011.

[66] D. Huron, *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2008.

[67] G. D. Webster and C. G. Weir, "Emotional responses to music: Interactive effects of mode, texture, and tempo," *Motivation and Emotion*, vol. 29, pp. 19–39, 2005.

[68] P. N. Juslin, "What does music express? basic emotions and beyond," *Frontiers in psychology*, vol. 4, p. 596, 2013.

[69] E. Schubert, "Modeling perceived emotion with continuous musical features," *Music perception*, vol. 21, no. 4, pp. 561–585, 2004.

[70] R. M. Holladay, A. D. Dragan, and S. S. Srinivasa, "Legible robot pointing," in *The 23rd IEEE International Symposium on robot and human interactive communication*. IEEE, 2014, pp. 217–223.

[71] A. Dragan and S. Srinivasa, "Generating legible motion," 2018.

[72] B. De Gelder, A. W. de Borst, and R. Watson, "The perception of emotion in body expressions," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 6, no. 2, pp. 149–158, 2015.

[73] P. Lietz, "Research into questionnaire design: A summary of the literature," *International journal of market research*, vol. 52, no. 2, pp. 249–272, 2010.

[74] C. L. Hardy and M. Van Vugt, "Nice guys finish first: The competitive altruism hypothesis," *Personality and Social Psychology Bulletin*, vol. 32, no. 10, pp. 1402–1413, 2006.

[75] P. Sheeran, "Intention—behavior relations: a conceptual and empirical review," *European review of social psychology*, vol. 12, no. 1, pp. 1–36, 2002.

[76] A. C. Horstmann and N. C. Krämer, "Great expectations? relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment," *Frontiers in psychology*, vol. 10, pp. 1–14, 2019.

[77] S. J. Czaja, N. Charness, A. D. Fisk, C. Hertzog, S. N. Nair, W. A. Rogers, and J. Sharit, "Factors predicting the use of technology: findings from the center

for research and education on aging and technology enhancement (create)." *Psychology and aging*, vol. 21, no. 2, pp. 333–352, 2006.

[78] C. McGinn and I. Torre, "Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 211–221.

[79] B. Kühnlenz, S. Sosnowski, M. Buß, D. Wollherr, K. Kühnlenz, and M. Buss, "Increasing helpfulness towards a robot by emotional adaption to the user," *International Journal of Social Robotics*, vol. 5, no. 4, pp. 457–476, 2013.

[80] J. A. Krosnick and D. F. Alwin, "An evaluation of a cognitive theory of response-order effects in survey measurement," *Public opinion quarterly*, vol. 51, no. 2, pp. 201–219, 1987.

Appendix A

Scenario Outlines for Sound Design

Introduction

It is 2030. In the past decade, the shortage of healthcare staff resulted in a growth in the usage of robots in hospitals. These are able to take all kinds of manual labor tasks over for which no human workers could be found. Luckily, robots have been further developed to take over the tasks of healthcare workers to accommodate to these problems. Also, the hospital in Enschede benefits from these advancements. Recently, they bought a number of robots (see pictures) that can take over transportation tasks. The robots carry around materials for on-demand deliveries as well as test samples such as blood samples. The robot is moving fully autonomously through the hospital and can load and unload all items by itself. By working completely on its own, the robot is a real relief for healthcare workers. In the weeks that it has been implemented, everyone felt the stress getting off their shoulders and they could focus on tasks that really require the knowledge and experience of the hospital staff and that is also fun to do for them. Nonetheless, for everyone, healthcare workers, patients, and visitors, seeing the robots is a rather new experience. Therefore, the robot is designed to work well in the human-centered environment to avoid the need of adjusting the hospital environment itself.

Warm-up: Robot entering the elevator with people in it

The robot is driving around the hospital to deliver the requested materials to a ward. The robot has to pick it up in the material storage room on the ground floor. After picking up some materials and starting its way to the ward the request came from, the robot is meeting a group of nurses passing by. The nurses seem to be of different age groups and have different experiences of working in the hospital. They pass the robot and smile at it. The robot continues its journey and arrives at the elevator. The ward that requested the material is on a different floor. The robot signals to the elevator that it wants to enter and after a few minutes, the elevator opens its doors. The robot wants to enter the elevator but there are already people in the elevator.

The robot wants to greet them when entering the elevator.

What sound does the robot make to get the intention of greeting people across in an emotional way so that it might trigger empathy?

Scenario 1: People blocking the way of the robot

For the robot, it is a busy day at the hospital. All day long, it is carrying around different items to deliver them to various locations. Currently, the robot is loaded with some materials the nurses need in the emergency room. On its way, the robot sees some kids playing in the entrance hall while their mother is talking at the reception. The robot goes into the hallway toward the emergency rooms. The robot goes around a corner but suddenly it stops. A group of nurses, doctors, and an older patient are standing in the hallway talking such that the robot cannot go around them. In the first moment, no one in the group realizes that they are standing in the way of the robot. But then the older woman notices the robot standing next to them and looks at the robot skeptically. It seems like she has never seen a robot before. Still, the rest of the group is so deep in their conversation, that they do not notice the robot right next to them. The robot urgently needs to deliver the materials and therefore needs the people to go out of the way.

What sound does the robot make to get the intention that the people need to move out of the way across in an emotional way so that it might trigger empathy?

cenario 2: A door on the way of the robot is closed

Later that week, the robot is driving through the hallways again to fulfill its tasks. It is on its way to the laboratory, to redistribute test samples to they can be analyzed. Currently, it is not busy in the hospital and the robot can easily make its way. The robot has to go around a young pregnant woman who is going slowly through the hospital with her partner who is talking on the phone. They stop for a second and look surprised at the robot. Also, some nurses are walking by the robot fast without really noticing it or giving it any attention. They seem to be in a hurry to get to a patient that needs their help. The robot is approaching a door. Normally, all doors on the main routes of the robot should be open. However, someone seems to have accidentally closed the door. When the robot arrives at the door it stops. It cannot open the door and therefore needs someone to open the door for it. It does not understand how this could happen. It wants to fulfill its tasks and therefore needs to get through the door somehow. Behind the robot, it hears steps and realizes that people are approaching it. These people can help the robot go through the door. When the people are going through the door, they need to hold it open for it, so that it can continue its journey.

What sound does the robot make to get the intention that the people need to open the door across in an emotional way so that it might trigger empathy?

Scenario 3: Robot getting stuck on an obstacle

Another day in that week, the robot is on its way to the material room to pick up some materials a nurse has demanded. The robot is going through the transportation corridors. The robot is approaching a cable that is laying on the ground. Normally, the sensors of the robot should detect the cable. However, today something must be wrong with the sensors and they do not detect it. Therefore, the robot is continuing to drive as always, but it is not built to be able to drive over cables. The wheels are too small to get over such obstacles. Hence, when trying to drive over the cables the robot gets stuck. One of the wheels made it still over the cable, but the robot is unable to get any further by itself. However, it has tasks to complete and wants to pick up the material to deliver them as soon as possible. The robot recognizes the issue it is having and stops trying to go over the cable. A few minutes later, it hears people approaching it. A nurse together with someone from the administrative staff is walking down the corridor. Perhaps they can help the robot.

What sound does the robot make to get the intention that the robot is stuck and needs help across in an emotional way so that it might trigger empathy?

Scenario 4: Robot loses its navigation

On the next day, the robot is requested to deliver materials to a ward that it seldomly goes to. The robot rarely needs to deliver material to this ward and cannot navigate further than this ward. The robot is making its way toward the ward. The robot wants to turn right around a corner but detects that people are coming from that hallway. A nurse is pushing an older man that seems to not be fully present in the moment through the hallway. To go out of the way of the people, it is going further straight instead of turning right. It does so to not be in the way of the nurse with the man. However, the robot has never gone further straight instead of turning left or right at that corner. Therefore, the robot loses its navigation and no longer knows where it is. As soon as the robot notices this, it stops. The robot cannot navigate back to the right hallway where it wanted to go or to the location before it tried to turn right. A family, the parents with a little boy and a slightly older girl, is approaching the robot. The robot recognizes its chance to get help.

What sound does the robot make to get the intention that the robot lost its navigation and needs help across in an emotional way so that it might trigger empathy?

Appendix B

Video Scripts

Scenario 1: People blocking the way of the robot
See long hallway that looks like a hallway with patient rooms. In the hallway, a nurse and a doctor are standing with a patient and their relatives. They are standing near the crossing of two hallways.

The group is blocking the hallway to go straight. They are the only people in the hallway.

The robot comes into the picture from behind the camera and is driving toward the group.

The robot gets closer and closer to the group. No one of the group seems to take notice of the approaching robot.

The robot is stopping behind the person that stands closest to it. The robot is stopping with a minimum distance of around 1 meter. The robot faces towards the group. There is space to go into the hallway to the right, however, the robot wants to go straight.

The group is still talking and still not seem to notice the robot.

The robot makes the sound and waits for a reaction from the group.

Video stops

Scenario 2: A door on the way of the robot is closed
See long hallway that looks like a hallway with patient rooms. In a bit distance, one can see an older couple sitting on chairs in front of a room. They look like they are waiting.

A doctor is walking quickly through the hallway together with a nurse towards the camera. They are talking.

The robot comes into the picture from behind the camera. It goes straight for a moment.

The robot slightly turns right. It seems like it wants to go in the nurse room on the right. The door of the nurse room is closed.

The robot stops in front of the door and positions itself next to the door.

From behind the robot (behind the camera) someone is approaching. The steps that the person makes can be heard.

The person appears in the frame and is walking towards the robot. The person is looking at the robot but seems like wanting to walk straight.

The person comes closer to the robot. When the person is in audible distance of the robot, it makes the sound and waits for a reaction from the person.

Video stops

Scenario 3: Robot getting stuck on an obstacle

See long hallway that looks like a hallway with patient rooms. In a bit distance, one can see an older couple talking to a doctor and nurse.

In the near distance a cable is lying on the ground.

The robot comes towards the camera. It goes straight towards the cable. It is lying on the ground in such a way that the robot cannot go around it.

From the opposite of the robot (behind the camera), someone is walking towards the robot. Their steps are slightly hearable.

The robot tries to drive over the cable but gets stuck. By driving back and forth it tries to get off the cable, but the robot is not able to do so.

After a few tries, the robot stops and stands still.

The person walking towards the robot is looking at the robot and observing the robot getting stuck. The robot looks at the person.

The person gets near the robot and is about to pass by. When the person is in audible distance, it makes the sound and waits for a reaction from the person.

Video stops

Scenario 4: The robot loses its navigation

See long hallway that looks like a hallway with patient rooms near a crossing of hallways. Voices of people are hearable.

The robot comes from the opposite of the camera towards the crossing.

The robot approaches a corner where it wants to turn right.

When arriving at the corner, the robot turns right and wants to go into the hallway.

In that moment, people arrive at that corner as well from the hallway the robot wanted to go in. The people are talking. To make space, the robot turns again by 90 degree and goes straight.

When being in the other hallway the robot stops when being in the hallway. A person is walking towards the robot (from behind the camera).

The person is approaching the robot. When the person is in an audible distance from the robot, it makes the sound and waits for a reaction from the person.

Video stops

Appendix C

Coding Scheme for Legibility

The answer does not relate to the problem of the robot or not identifying the problem of the robot (0):

- The robot is greeting
- The robot is trying to say hello
- The robot makes music
- The robot is patiently waiting
- The robot wants to talk to the group
- Description of “wrong” problem
- The robot acknowledges people (anything) around them

The answer is a general description of the situation (1):

- The robot is stuck
- The robot needs help
- The robot has a problem
- The robot has an error
- The robot is malfunctioning
- The robot’s path is blocked
- It cannot continue its path

The answer is a specific description of what the situation is or a combined description (2):

- The robot is stuck and needs help
- Context 1
  - People are blocking the way
  - People should move out of the way
  - The robot needs to go where the people are standing
- Context 2
  - The robot needs the door to be opened
  - Wants to go through the closed door
- Context 3
  - The robot needs to get over the extension cord
  - The robot cannot move over the cord

Appendix D

Coding Scheme for Emotions

Neutral Emotions (0):

- No emotion
- Normal
- Neutral
- Any answer that does not describe an emotion

Positive Emotions (1):

- Patience
- Happy
- Politeness
- Inquisitive
- Playful
- Pleasant
- Intrigued
- Cheerful
- Content
- Caring
- Delight

Negative Emotions (2):

- Annoyance
- Concern
- Disturbed
- Distress
- Frustration
- Alert
- Sadness
- Confusion
- Urgency
- Impatience
- Helpless
- Unsurity
- Disgust
- Troubled
- Alert
- Disappointed
- Disgust
- Disgruntled
- Apprehensive
- Troubled
- Despair
- Conflicted
- Anxious
- Irritation

Appendix E

Statements to Measure Empathy

Statements with (CE) behind them measure cognitive empathy. Statements with (AE) behind them measure affective empathy. The original order of the statements was as below. However, all participants saw the statements in a random order

I could easily tell the feelings of the robot. (CE)

It was easy to understand the emotions of the robot. (CE)

The robot influenced how I am feeling. (AE)

It was easy to imagine how the robot was feeling. (CE)

I stayed emotionally detached from the situation. (AE)

I got deeply involved with the feelings of the robot. (AE)

The situation of the robot affected me very much. (AE)

I considered the feelings of the robot. (CE)

I find it easy to put myself in the shoes of the robot. (CE)

Appendix F

Results of Factor Analysis for Empathy

Table F.1: Factor Loadings for Scenario 1.

	Factor 1	Factor 2
Question 1	.85	.32
Question 2	.88	.32
Question 3	.27	.75
Question 4	.82	.32
Question 5	.19	.30
Question 6	.33	.79
Question 7	.22	.83
Question 8	.41	.71
Question 9	.64	.34

Table F.2: Factor Loadings for Scenario 2.

	Factor 1	Factor 2
Question 1	.88	.26
Question 2	.88	.24
Question 3	.25	.88
Question 4	.90	.25
Question 5	.14	.37
Question 6	.34	.78
Question 7	.23	.89
Question 8	.63	.56
Question 9	.67	.29

Table F.3: Factor Loadings for Scenario 3.

	Factor 1	Factor 2
Question 1	.89	.26
Question 2	.90	.24
Question 3	.21	.88
Question 4	.93	.23
Question 5	.15	.42
Question 6	.32	.85
Question 7	.23	.84
Question 8	.57	.59
Question 9	.75	.31

Table F.4: Factor Loadings for All Scenarios Combined Excluding Question 5 and 8.

	Factor 1	Factor 2
Question 1	.87	.28
Question 2	.89	.27
Question 3	.24	.83
Question 4	.89	.27
Question 6	.33	.79
Question 7	.22	.86
Question 9	.68	.31

Appendix G

Coding Scheme for Prosocial Behavior

Prosocial behavior is any behavior that merely benefits someone else (in this case the robot) and does not have a direct added value for the person executing it

No prosocial behavior (0):

- The behavior is not directly to the benefit of the robot
- No prosocial behavior is described
- Behavior that might be prosocial but does not relate to resolving the situation for the robot

Weak prosocial behavior (1):

- Prosocial behavior that only indirectly benefits the robot, only indirectly resolves the problem/issue
- Try to help the robot
- Think about how the robot can be helped
- See what is going on or what the robot wants
- Interacting with the robot
- Whether they would engage in prosocial behavior or not depends on something
- Stop and see what is happening
- I would move out of the way
- Make sure things are okay

Strong prosocial behavior (2):

- Prosocial behavior that directly resolves the problem/issue of the robot
- Tell someone the robot is stuck
- Look for someone who can help
- Seeking advice
- Context 1: Ask the rest of the group to move out of the way
- Context 2
  - Open the door for the robot
  - Knock the door
- Context 3: Move the cable out of the way

Appendix H

Questionnaire

Evaluating a Hospital Robot for Transportation Tasks

Study Information

Please read the study information below carefully!

DESCRIPTION: You are invited to participate in a survey to evaluate the behavior of a hospital robot. The robot is developed to take over transportation tasks of test samples and materials needed in the wards of the hospital. The robot will navigate and drive through the hospital. You will see three videos of situations with the robot driving through the hospital to pick up material. You will be asked to evaluate the situations and describe your impressions of the robot through a number of open and closed questions concerning your interpretation of the behavior of the robot. We are interested in your personal evaluation and therefore there are no right or wrong answers to the questions. Try to go along with the initial thoughts and answers you have.

TIME INVOLVEMENT: Your participation will take around 10-15 minutes.

RISKS AND BENEFITS: To the best of our knowledge, there is no discomfort involved in this study, nor any risks. Every participant will receive a financial reward of 3$. Besides, we cannot and do not guarantee or promise that you will receive any benefits from this study. This research was reviewed by the Ethics Committee of Computer and Information Sciences of the University of Twente.

PARTICIPANT'S RIGHTS: Please understand that your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time before or during the survey without giving any reasons. You have the right to refuse afterward (within 24 hours) to allow your data to be used for the research.

The results of this research study will be summarized in a report that will be made available online. Your identity will not be made known in written materials resulting from the study. All your data will be made anonymous at the earliest possible stage. Research data will be stored in such a way that unauthorized access will be minimized.

CONTACT INFORMATION

Questions: If you have any questions, concerns, or complaints about this research, its procedures, risks, and benefits, contact the researcher, Theresa Höfker, through Prolific or by email t.hofker@student.utwente.nl.

You can also contact the supervisors of the researcher Bob Schadenberg at +31534891011 or by email b.r.schadenberg@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H238 and Khiet Truong at +31534893683 or by email k.p.truong@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H217.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Ethics Committee, Faculty of EEMCS, University of Twente, +31534896719, ethicscommittee-cis@utwente.nl

Next

Evaluating a Hospital Robot for Transportation Tasks

Informed Consent

Please tick the applying boxes below

Taking part in the study:

* 1. I have read and understood the study information.

◯ Yes

* 2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions by withdrawing from the study at any time, without having to give a reason.

◯ Yes

◯ No

* 3. I understand that taking part in the study involves answering open and closed questions regarding the understanding of the behavior of the robot.

◯ Yes

Use of information in the study

* 4. I understand that the information I provide will be analyzed and the results used for a report which will be shared with the supervisors of the project and based on the results a report will be created that will be shared online.

◯ Yes

* 5. I understand that through the survey platform Personal Identifiable Information, such as my IP address, possibly is collected. I understand that this data is removed at the earliest moment possible. I understand that personal information collected about me that can identify me, will not be shared beyond the study team.

◯ Yes

* 6. I agree that my information can be quoted in research outputs. All personal information will be anonymized that it cannot be traced back to me as an individual.

◯ Yes

◯ No

* 7. I give permission that the data I provide in the study can be used for future research and learning. I understand that the data will be archived fully anonymized so that it cannot be traced back to me as an individual.

◯ Yes

◯ No

Study contact details for further information:

If you have any questions, concerns, or complaints about this research, its procedures, risks, and benefits, contact the researcher, Theresa Höfker, through Prolific.

You can also contact the supervisors of the researcher Bob Schadenberg at +31534891011 or by email b.r.schadenberg@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H238 and Khiet Truong at +31534893683 or by email k.p.truong@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H217.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Ethics Committee, Faculty of EEMCS, University of Twente, +31534896719, ethicscommittee-cis@utwente.nl

Next

Evaluating a Hospital Robot for Transportation Tasks

Informed Consent

Please tick the applying boxes below

Taking part in the study:

! This question requires an answer.

* 1. I have read and understood the study information.

◯ Yes

! This question requires an answer.

* 2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions by withdrawing from the study at any time, without having to give a reason.

◯ Yes

◯ No

! This question requires an answer.

* 3. I understand that taking part in the study involves answering open and closed questions regarding the understanding of the behavior of the robot.

◯ Yes

Use of information in the study

! This question requires an answer.

* 4. I understand that the information I provide will be analyzed and the results used for a report which will be shared with the supervisors of the project and based on the results a report will be created that will be shared online.

◯ Yes

! This question requires an answer.

* 5. I understand that through the survey platform Personal Identifiable Information, such as my IP address, possibly is collected. I understand that this data is removed at the earliest moment possible. I understand that personal information collected about me that can identify me, will not be shared beyond the study team.

◯ Yes

! This question requires an answer.

* 6. I agree that my information can be quoted in research outputs. All personal information will be anonymized that it cannot be traced back to me as an individual.

◯ Yes

◯ No

! This question requires an answer.

* 7. I give permission that the data I provide in the study can be used for future research and learning. I understand that the data will be archived fully anonymized so that it cannot be traced back to me as an individual.

◯ Yes

◯ No

Study contact details for further information:

If you have any questions, concerns, or complaints about this research, its procedures, risks, and benefits, contact the researcher, Theresa Höfker, through Prolific.

You can also contact the supervisors of the researcher Bob Schadenberg at +31534891011 or by email b.r.schadenberg@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H238 and Khiet Truong at +31534893683 or by email k.p.truong@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H217.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Ethics Committee, Faculty of EEMCS, University of Twente, +31534896719, ethicscommittee-cis@utwente.nl

Next

Evaluating a Hospital Robot for Transportation Tasks

Introduction

In this study, you will see a total of three short videos. In the videos, you will see a hospital hallway of patient rooms. A robot is driving through the hallway to pick up some materials. Please pay close attention to the videos and watch them carefully. Pay close attention to the robot and what it is doing. After you watched one video you will progress to a set of questions. These questions are about your impression of the robot. Importantly, there are no right or wrong answers and we are interested in your judgment. Try to go with your first initial response. At the end of the questionnaire, you will be asked some background questions. After these, the questionnaire ends.

* 8. What is your experience with robots?

	1 (No experience at all)	2	3	4	5	6	7 (Daily work/interaction with robots)
Experience	◯	◯	◯	◯	◯	◯	◯

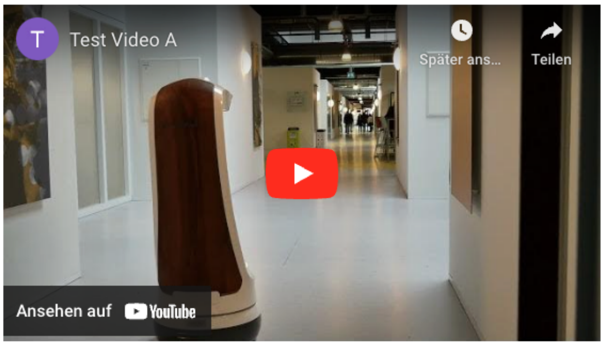
Next

Evaluating a Hospital Robot for Transportation Tasks

Test Video

You are going to watch three videos in which it is important that you watch and listen to what happens in the video. Therefore, please ensure that you have your **audio turned on.** Accordingly, please watch the test video below and ensure that you can clearly hear the sounds by adjusting your audio volume. Pay close attention to the robot, its behavior, and its communication. After watching, please answer the question below and proceed to the next page.

Test Video A



* 9. Which test video did you see? (You can find the name of the video above the video)

- ◯ Test Video A
- ◯ Test Video B
- ◯ Test Video C

Next

Evaluating a Hospital Robot for Transportation Tasks

Video A

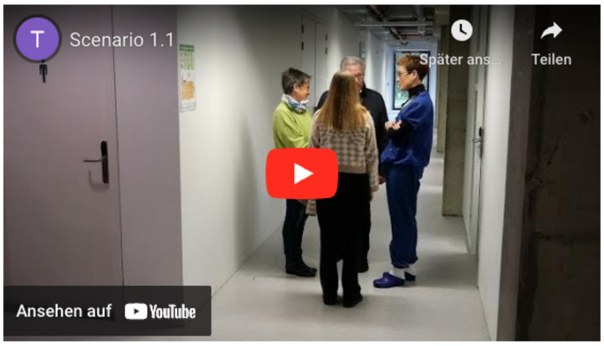
* 10. Please describe shortly what the robot sounded like in the previous video.

Next

Evaluating a Hospital Robot for Transportation Tasks

Scenario 1.1 Video

Below is a video of a robot driving through the hallway of a hospital to pick up some material to deliver. Please watch the video below **carefully** and pay close attention to the **robot** and its **behavior**. Only watch the video **once** and then go to the next page.



* 11. Did you watch the video until the end?

◯ Yes

Next

Evaluating a Hospital Robot for Transportation Tasks

Scenario 1.1

Please answer the questions below honestly. There are no right or wrong answers, the questions are about your impression and evaluation. Try to go with the first answer that comes to your mind

* 12. 1.1 In which direction is the robot looking at the end of the video?

- ◯ To the left
- ◯ To the group
- ◯ To the camera
- ◯ To the right

* 13. What is the robot trying to communicate at the end of the video?

[ ]

* 14. How confident are you in your answer to the previous question?

	1 (Not confident at all)	2	3	4	5	6	7 (Completely confident)
Confidence	◯	◯	◯	◯	◯	◯	◯

* 15. What emotion is the robot displaying? (Think about emotions like happy, sad, disgusted, neutral, angry)

[ ]

* 16. How strong do you feel the emotion is?

	1 (Not strong at all)	2	3	4	5	6	7 (Very Strong)
Emotion strength	◯	◯	◯	◯	◯	◯	◯

* 17. Please rate to what extent these statements are true.

	1 (Not true at all)	2	3	4	5	6	7 (Completely true)
I got deeply involved with the feelings of the robot.	○	○	○	○	○	○	○
The situation of the robot affected me very much.	○	○	○	○	○	○	○
It was easy to imagine how the robot was feeling.	○	○	○	○	○	○	○
I considered the feelings of the robot.	○	○	○	○	○	○	○
I find it easy to put myself in the shoes of the robot.	○	○	○	○	○	○	○
It was easy to understand the emotions of the robot.	○	○	○	○	○	○	○
The robot influenced how I am feeling.	○	○	○	○	○	○	○
I could easily tell the feelings of the robot.	○	○	○	○	○	○	○
I stayed emotionally detached from the situation.	○	○	○	○	○	○	○

* 18. What would you do if you are one of the persons in the group in the situation?

Next

Evaluating a Hospital Robot for Transportation Tasks

Background Questions

* 35. What is your age?

* 36. What is your gender?

- ◯ Female
- ◯ Male
- ◯ Non-binary
- ◯ Rather not say
- ◯ Other (please specify)

Next

Evaluating a Hospital Robot for Transportation Tasks

Debriefing

This is the end of the survey. Thank you for filling out the survey. Your participation is highly valued! Please read the debriefing, fill out the questions and then click done.

In the beginning, only general information about the purpose of the study was given and the exact purpose of the study was not explained to avoid influencing you. Therefore, you can find more detailed information about the purpose of the study below.

The study is concerned with whether people can understand robots better when they make use of musical sequences to communicate their intentions and emotions and how people intend to respond. In total there were three types of sounds, namely musical sounds, beeping sounds, and no sounds of which you heard one in the videos. We measure how easy people find it to understand and interpret the behavior and especially the sounds of the robot. Also, it is measured whether people associate emotions with the robot. Furthermore, we are looking at whether people feel empathetic with the robot. Lastly, we measure whether people intend to help the robot. The answers to these variables are compared between the conditions and we will look at whether there are differences between the conditions.

* 37. After having received more insights into the purpose of the research, do you still consent that your data will be used for research purposes?

◯ Yes

◯ No

Prolific

* 38. What is your Prolific ID?

[ ]

Please copy and paste the following code as the **completion code** in Prolific: C605PP8Q
Use this code to **register your completion** on Prolific.

* 39. Have you entered the code?

◯ Yes

If you have any further questions, you can contact the researcher (Theresa Höfker) through Prolific or by email t.hofker@student.utwente.nl.

You can also contact the supervisors of the researcher Bob Schadenberg at +31534891011 or by email b.r.schadenberg@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H238 and Khiet Truong at +31534893683 or by email k.p.truong@utwente.nl, address: University of Twente, 7500 AE Enschede, Netherlands. Building: Citadel, Room: H217.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Ethics Committee, Faculty of EEMCS, University of Twente, +31534896719, ethicscommittee-cis@utwente.nl

Thank you again for your contribution. You are of great help to me to complete my Master's Thesis.

Done

Appendix I

Model Fit Measures for Legibility

Table I.1: Model fit measures for legibility and ANOVA results.

Model	df	AIC	BIC	Log-Likelihood	$\chi 2$	p
(1) Sound	5	501.97	519.28	-246.98	-	-
Model 1 with covariate						
(2) Age	6	501.04	522.69	-245.52	2.93	.087
(3) Experience	6	503.82	525.47	-246.91	0.14	.705

Note. Model 2, 3, 4 were compared to model 1, Model 5 was compared to model 2

Table I.2: Model fit measures for confidence in legibility and ANOVA results.

Model	df	AIC	BIC	Log-Likelihood	$\chi 2$	p
(1) Sound	5	2117.46	2139.13	-1053.74	-	-
Model 1 with covariate						
(2) Age	6	2103.04	2129.02	-1045.52	16.44	.000
(3) Gender	7	2116.99	2147.30	-1051.49	4.49	.105
(4) Experience	6	2111.48	2137.46	-1049.74	8.00	.000
Model 2 with a second covariate						
(5) Age and experience	7	2092.69	2123.00	-1039.35	12.35	.000

Note. Model 2, 3, 4 were compared to model 1, Model 5 was compared to model 2

Appendix J

Test of Assumptions for Confidence in Communication Interpretation

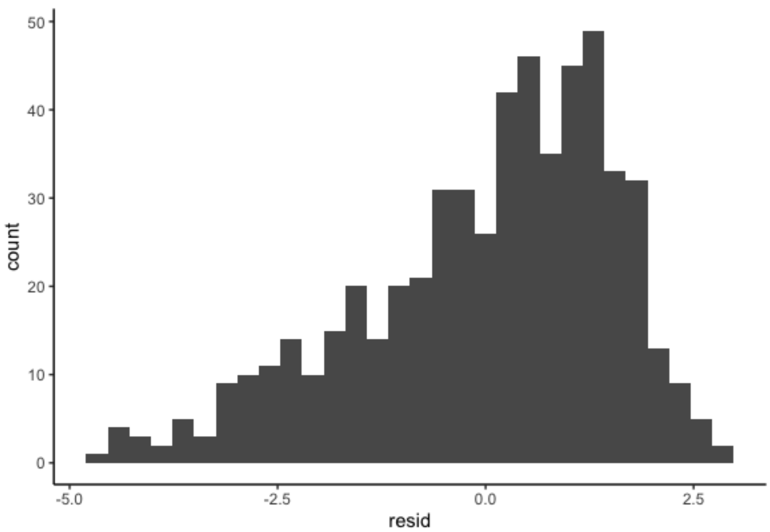


Figure J.1: Histogram of distribution of residuals for the confidence in the interpretation of the robot communication.

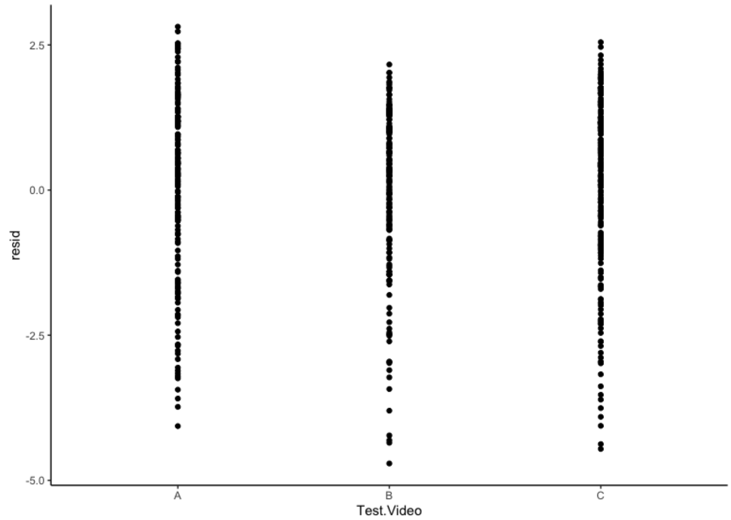


Figure J.2: Scatterplot of residuals by sound condition for the confidence in the interpretation of the robot communication.

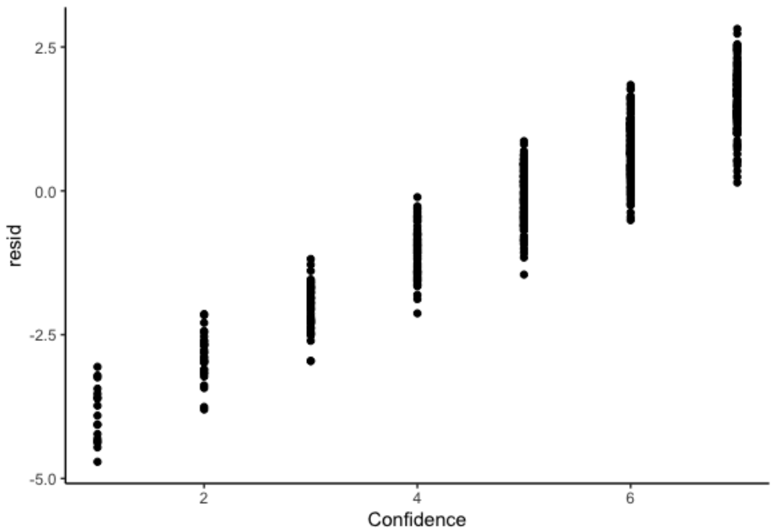


Figure J.3: Scatterplot of residuals by confidence for the confidence in the interpretation of the robot communication.

Appendix K

Model Fit Measures for Emotions

Table K.1: Model fit measures for perceived emotion and ANOVA results.

Model	df	AIC	BIC	Log-Likelihood	$\chi2$	p
(1) Sound	5	973.61	995.26	-481.80	-	-
Model 1 with covariate						
(2) Age	6	971.52	997.49	-479.76	4.09	.043
(3) Experience	6	971.22	997.20	-479.61	4.38	.036
Model 3 with a second covariate						
(4) Age and experience	7	970.20	1000.51	-478.10	3.02	.082

Note. Model 2, 3, 4 were compared to model 1, Model 5 was compared to model 2

Table K.2: Model fit measures for perceived emotion strength and ANOVA results.

Model	df	AIC	BIC	Log-Likelihood	$\chi2$	p
(1) Sound	5	2259.41	2281.03	-1123.71	-	-
Model 1 with covariate						
(2) Age	6	2266.76	2292.69	-1127.38	5.34	.021
(3) Gender	7	2257.81	2288.05	-1121.90	5.61	.061
(4) Experience	6	2263.64	2289.58	-1125.82	2.22	.136

Note. Model 2, 3, 4 were compared to model 1, Model 5 was compared to model 2

Appendix L

Test of Assumptions for Perceived Strength of Emotion

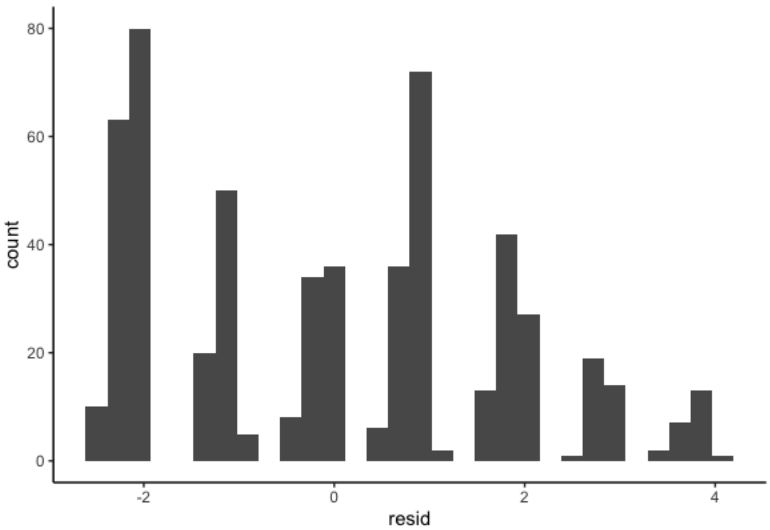


Figure L.1: Histogram of distribution of residuals for perceived strength of emotion.

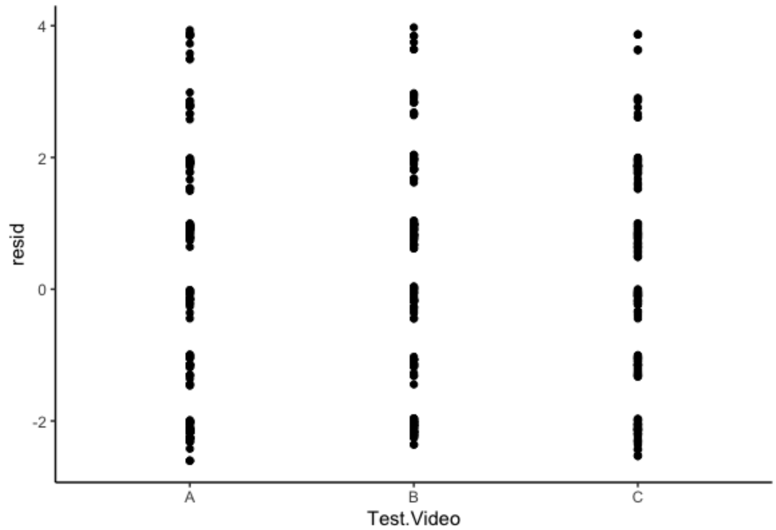


Figure L.2: Scatterplot of residuals by sound condition for perceived strength of emotion.

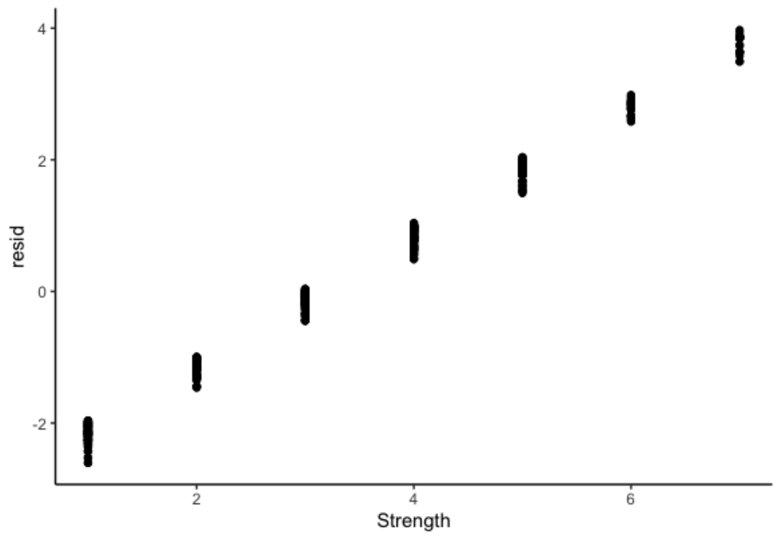


Figure L.3: Scatterplot of residuals by confidence for perceived strength of emotion.

Appendix M

Model Fit Measures for Empathy

Table M.1: Model fit measures for cognitive empathy and ANOVA results.

Model	df	AIC	BIC	Log-Likelihood	$\chi2$	p
(1) Sound	5	2187.50	2209.12	-1088.75	-	-
Model 1 with covariate						
(2) Age	6	2196.93	2222.87	-1092.47	7.43	.006
(3) Gender	7	2191.45	2221.70	-1088.73	0.05	.977
(4) Experience	6	2192.79	2218.72	-1090.39	3.29	.069

Table M.2: Model fit measures for affective empathy and ANOVA results.

Model	df	AIC	BIC	Log-Likelihood	$\chi2$	p
(1) Sound	5	1905.42	1927.04	-947.71	-	-
Model 1 with covariate						
(2) Age	6	1915.05	1940.98	-951.52	7.63	.005
(3) Gender	7	1910.22	1940.47	-948.11	0.80	.669
(4) Experience	6	1909.60	1935.54	-948.80	2.18	.140

Appendix N

Test of Assumptions for Affective Empathy

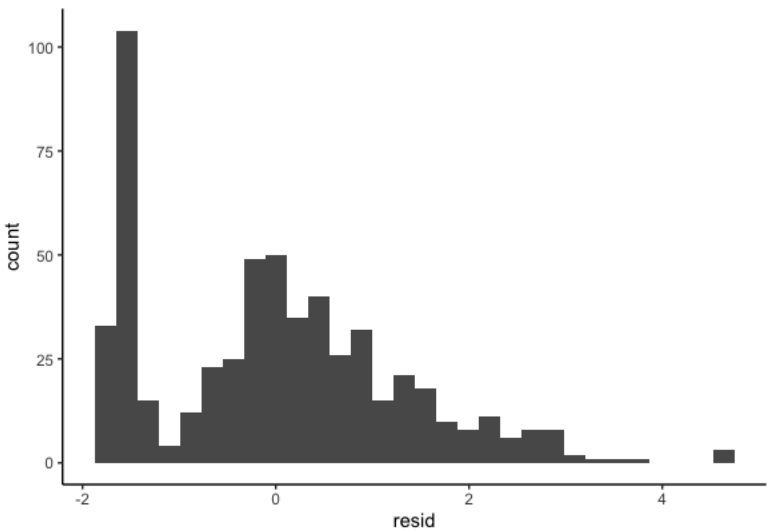


Figure N.1: Histogram of distribution of residuals for affective empathy.

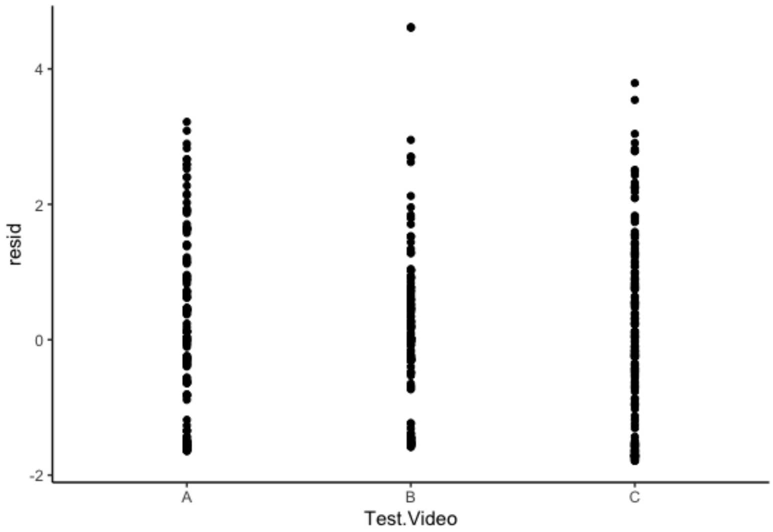


Figure N.2: Scatterplot of residuals by sound condition for affective empathy.

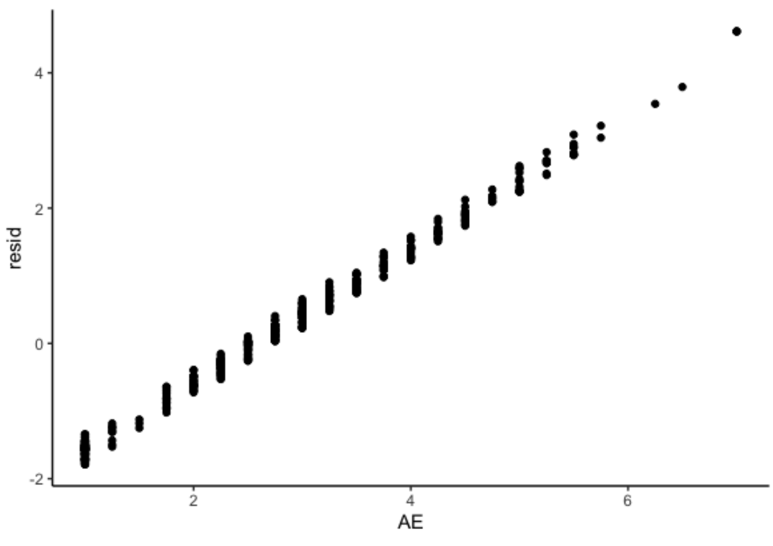


Figure N.3: Scatterplot of residuals by confidence for affective empathy.

Appendix O

Test of Assumptions for Cognitive Empathy

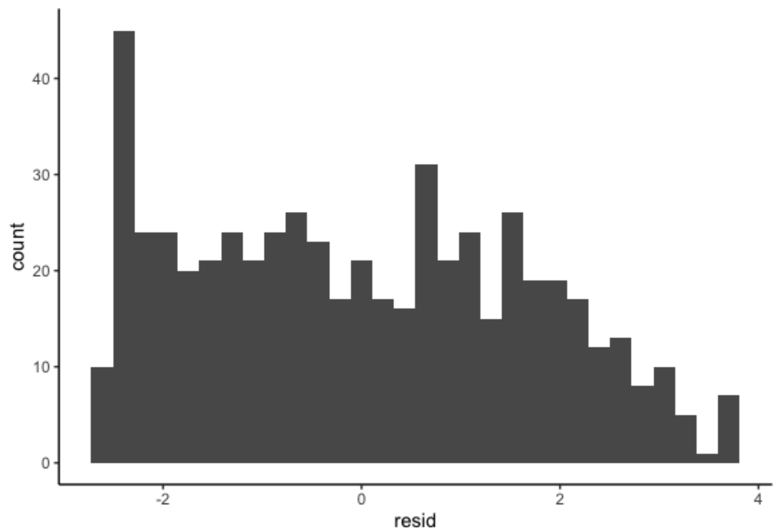


Figure O.1: Histogram of distribution of residuals for cognitive empathy.

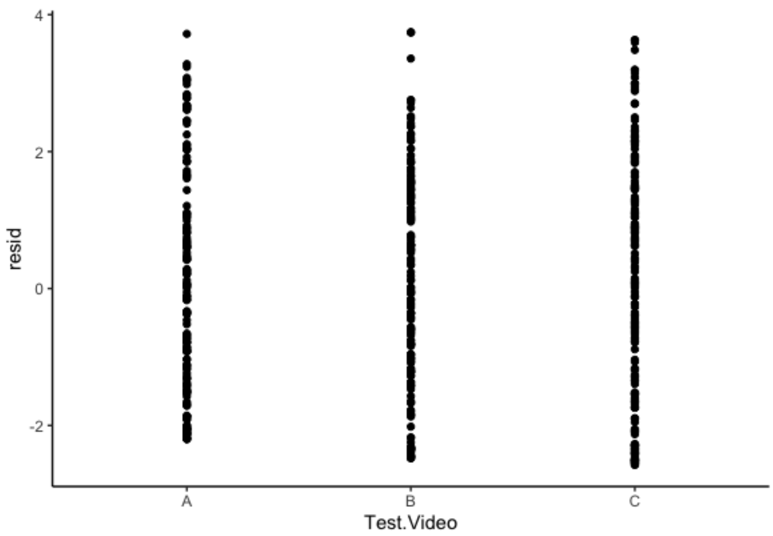


Figure O.2: Scatterplot of residuals by sound condition for cognitive empathy.

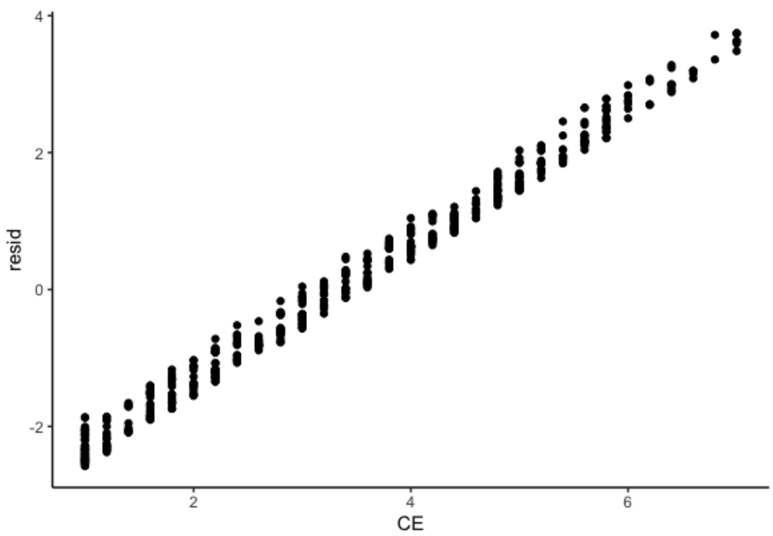


Figure O.3: Scatterplot of residuals by confidence for cognitive empathy.

Appendix P

Model Fit Measures for Prosocial Behavior

Table P.1: Model fit measures for prosocial behavior and ANOVA results.

Model	df	AIC	BIC	Log-Likelihood	$\chi 2$	p
(1) Sound	5	1135.57	1157.32	-562.83	-	-
Model 1 with covariate						
(2) Age	6	1137.58	1163.56	-562.79	0.09	.769
(3) Gender	7	1134.66	1164.97	-560.33	5.01	.081
(4) Experience	6	1136.30	1162.27	-562.15	1.38	.241