Toward Context-Aware, Affective, and Impactful Social Robots

By
Morten Roed Frederiksen

A dissertation presented for the degree of
Doctor of Philosophy

IT University
of Copenhagen
Denmark
September 30, 2021

Supervisor: Kasper Stoy (ksty@itu.dk)
Co-Supervisor: Jonas Fritsch (frit@itu.dk)

Acknowledgements

I would like to thank all the people who have supported me while conducting this research project.

Firstly, I would like to thank Kasper Støy at the IT-university of Copenhagen Denmark for giving me incredible amounts of support and the courage to create my own path in robotics research.

Secondly, I would like to thank my family for having unlimited patience with me throughout my time at the university, for giving me thoughtful input, and for proofreading the various robotics research publications I authored in academia.

Thirdly, I would like to thank both the REAL lab and AIR lab at the IT-University of Copenhagen. The skilled employees at those research groups provided valuable insights during the development of the robots used in this project.

1 Summary

When robots work in social contexts they are often required to partake in increasingly complex social interactions. This sets high requirements for their social capabilities. To participate in complex social interactions, the robots must be properly understood by the humans with whom they interact. This effect can be influenced by how the robots express themselves and how they are emotionally perceived. The more robots are able to comprehend whom they are interacting with and where the interaction occurs, the more they can adapt their behaviors to different scenarios. This adaption can potentially make it possible to optimize how they are perceived and as a result make them better equipped for handling complex social interactions.

We, humans express ourselves with both verbal and nonverbal communication methods to convey how we feel, and we do this with a deep understanding of the context of the interaction and the people we interact with. To us as humans, this task is often trivial when interacting in familiar environments and with people we know. When we no longer have the advantage of familiarity with known contexts, we often have to rely on our ability to interpret the cues and signals of the immediate situation. As a path towards improving the affective abilities of robots, it is crucial that we focus on the robots' ability to understand the immediate context. Even a simple understanding of the context will allow them to adapt to both contextual changes as well as to the humans they interact with. These abilities are vital for strengthening their affective impact and successfully introducing them into complex social scenarios. This dissertation reports the conduct and results of a series of experiments that aims to contribute to our understanding of how to create more impactful and context-aware affective robots. The subject was investigated through robot engineering and experiments in human-robot interactions.

Inspired by patterns in human-human interactions, the experimental setup outlined in this dissertation aimed at highlighting both the engineering and behavioral aspects of each human-robot interaction with the aim to strengthen the affective impact of social robots. In each of our interactions, we humans often express ourselves using multiple interaction modalities. These include the way we gesture, how we move, how we speak, and even how we look. To understand the specific synergy among these interaction modalities, we designed and constructed two non-humanoid robots. The technical and behavioral implementations of the robots were verified through multiple human-robot interactions, and the results contribute to the research field of affective robots in a number of ways. The initial finding defines a model for systematically assessing and characterizing the affective strengths of a robot. The model is useful both for the comparison of robots and as a guideline for roboticists in the design phase of affective robots. By applying the model to existing social robots it was also found that the use of multiple interaction modalities for robots in human-robot interactions has an untapped potential for success. The second finding outline how the coordination and specific timing of a robot's reactions in an interaction influence the robot's affective impact and alter how humans perceive it. That is, the project showed that when robots can respond to a broad variety of input types with the proper timing, humans perceive them as having a greater affective impact.

The current generation of social robots works fairly well in the context for which they have been designed. However, they rarely adapt their (affective) behavior to the changes in the environment, and they often struggle to comprehend the social requirements of their interaction. Throughout each of our interactions, we humans regulate our behaviors by interpreting subtle cues from the people we interact with and by understanding the constraints of the places in which we interact. This may be as simple as not laughing when someone is sad or lowering our voice in a smaller space. We also adapt to the mood of the people we interact with and adjust our behaviors to the requirements of the physical context. To navigate interactions that demand such contextual comprehension, robots need a subset of the same skills. In this project, we designed a humanoid robot to facilitate autonomous context awareness through human-robot interactions combined with context-informing cues in the physical environment. We also presented a system that enables robots to adapt to different physical contexts using immediately available sensor data in each interaction. We found that simple context awareness in robots can be facilitated using data that are easily attainable from the physical context. The system is applicable to other robots and requires only simple sensors available in most robots. Finally, we came up with a method that allows a robot to adapt to different users based on simple sensors and through a short-duration interaction. Through our experiments, we found

that the speech and movement patterns of humans gained from the initial moments of an interaction could sufficiently be used to distinguish individual users and could potentially facilitate further user adaptation of robot behaviors.

This dissertation argues for the need to shift our perspective on and approach to designing, constructing, and testing social robots, with the aim of increasing the affective impact of robots. The approach focuses on simultaneously using multiple simple interaction modalities to optimize how robots convey complex affective information. This is in contrast to using a single but highly specialized interaction modality. The approach also focuses on combining several sources of context information to gain knowledge on the circumstances of the interaction, and on adapting to such circumstances to create more impactful and believable robots. It aims to outline how simple, information obtainable from the immediate interactions between humans and robots can help the robots become more context aware and have a stronger affective impact. The simple information in our experiments consisted of the physical dimensions of the test environment while the measured human attributes consisted of the speech and movement characteristics of each participant. Using such information may give robots the ability to adapt to changes in the physical context and to meet the user-specific behavioral demands of each interaction.

As a future strategy, this dissertation suggests that robot designers change their perspective on when to use contextual knowledge and decrease the requirements on systems that provide contextual comprehension. Although a complete and human-like understanding of the current context may not be possible with the current technology, it is beneficial to already use the available contextual information in robots, as even simplistic context information may be useful for informing affective robot behaviors.

2 Summary (In Danish)

Når robotter skal arbejde i sociale sammenhænge, er det ofte påkrævet at de tager del i komplekse sociale interaktioner. Dette kan være udfordrende for deres sociale egenskaber. For at kunne deltage i sådanne interaktioner kræves det, at robotterne kan udtrykke sig på en måde, der er tydelig og forståelig for de mennesker, de interagerer med. Den forståelse kan påvirkes af måden hvorpå robotter udtrykker sig i interaktioner, samt hvordan de emotionelt bliver opfattet af mennesker. Jo bedre robotter er i stand til at forstå de mennesker de interagerer med og forstå de fysiske omstændigheder, hvori de interagerer, desto bedre kan de tilpasse sig deres omgivelser. Denne tilpasning kan potentielt set muliggøre, at robotter bliver i stand til forbedre, hvordan de bliver opfattet af mennesker i forskellige situationer. Dermed bliver de bedre udrustet til at håndtere komplekse sociale interaktioner.

Vi mennesker bruger både verbale og non-verbale kommunikationsmetoder til at udtrykke, hvordan vi har det. Dette gør vi med en dyb forståelse for den aktuelle kontekst, samt for de mennesker vi interagerer med. For os mennesker er dette ofte en let opgave, når vi omgåes de mennesker, vi kender, og vi befinder os i omgivelser, vi er vante til. Når vi derimod er i uvante omgivelser, må vi forsøge at tilpasse os de ændrede fysiske og sociale kontekster. Dette gør vi ved at forstå adfærdsmæssige signaler fra de mennesker, vi interagerer med og ved at forstå simple tegn fra de fysiske omgivelser.

Robotter har brug for affektive egenskaber for at kunne kommunikere følelsesmææssige informationer til mennesker. Det giver derfor mening at forbedre netop de affektive egenskaber hos robotter med det formål at gøre dem i stand til at håndtere komplekse sociale interaktioner. En mulig forbedring af disse færdigheder kan nåes gennem robotters umiddelbare forståelse af deres kontekst. Selv en simpel forståelse af konteksten kan muliggøre en tilpasning af deres opførsel til eventuelle ændringer i både de fysiske

omstændigheder samt hos de mennesker, med hvem de interagerer. Denne afhandling præsenterer indholdet og resultaterne af en serie af eksperimenter, der har til formål at skabe bedre affektive og kontekstforstående robotter. Disse områder er blevet undersøgt gennem robotudvikling samt eksperimenter i interaktioner mellem mennesker og robotter.

Ved at se på de mønstre, som mennesker interagerer med hinanden i, har dette projekt forsøgt at afdække både de konstruktionsmæssige og adfærdsmæssige aspekter af mennesker-robot interaktioner for at styrke sociale robotters affektive indvirkning. I hver interaktion udtrykker vi mennesker os ved hjælp af alle tilgængelige kommunikationskanaler. Dette inkluderer den måde hvormed vi gestikulerer, hvordan vi bevæger os, hvordan vi taler og endda hvordan vi ser ud. Alle disse kommunikationskanaler hjælper os til at udtrykke os klart og tydeligt. For at forstå synergien mellem disse interaktionsmodaliteter, designede og konstruerede vi to ikke-humanoide robotter.

Vi testede robotternes tekniske og adfærdsmæssige egenskaber i flere interaktioner mellem mennesker og robotter, og resultaterne heraf bidrager til den generelle viden om affektive robotter på følgende måder: Det første projekt definerer en model til systematisk at kunne vurdere og karakterisere en robots affektive styrker. Modellen kan bruges både som et sammenligningsgrundlag imellem robotter samt som en ledesnor for robotdesignere i designfasen af affektive robotter. Ved at bruge modellen på et udsnit af aktuelle affektive robotter kan vi se, at der er et ubenyttet potentiale i at kombinere flere interaktionsmodaliteter, når affektive robotter kommunikerer. Resultaterne fra vores andet projekt beskriver, hvordan den specifikke timing af robotters reaktioner i en interaktion påvirker, hvordan mennesker opfatter robotten. Resultaterne indikerer, at når robotter reagerer på en bred vifte af input-typer med en velovervejet timing, så opfatter mennesker dem som havende en større affektiv indvirkning på interaktionen.

Mange nuværende sociale robotter fungerer godt i den sammenhæng, de er designet til at arbejde i. Det er dog ofte, at de bliver udfordret, når miljøet omkring dem ændrer sig. De er også udfordret i at kunne tilpasse sig, når de sociale kontekstuelle krav ændrer sig i interaktioner med mennesker.

Når vi mennesker interagerer med andre mennesker, så tilpasser vi vores adfærd til hinanden. Dette gør vi ud fra en tolkning af simple signaler fra de mennesker, vi interagerer med. Vi gør det også ved at forstå de fysiske omstændigheder for vores interaktion. Det kan være helt simple ting, vi tilpasser, som for eksempel at vi undlader at grine, når vi snakker med en person, der er ked af det, eller at vi dæmper stemmen, når vi er tæt på andre mennesker. For at kunne håndtere komplekse sociale interaktioner har robotterne brug for lignende færdigheder. I dette projekt har vi designet en humanoid robot for at undersøge, hvordan vi kan facilitere kontekstbevidsthed for robotter i interaktioner med mennesker. Dette opnår robotten gennem en forståelse af simple kontekstuelle informationer ud fra de fysiske omstændigheder for interaktionen. Vores resultater indikerer, at en simpel forståelse af den fysiske kontekst kan opnåes med let tilgængelig data fra den umiddelbare interaktion. Systemet er direkte anvendeligt for andre robotter og har begrænsede hardwaremæssige krav. Systemet benytter udelukkende sensorer, der er til stede i de fleste robottyper. Vi designede en metode, der muliggører, at en robot tilpasser sig forskellige brugere. Dette var også baseret på simple sensorer og krævede kun korte interaktioner for at opsamle tilstrækkeligt data. Vi kom frem til, at talemønstre og bevægelsesmønstre kunne bruges til at genkende individuelle brugere i enkelte kontekster, og at disse data potentielt kan bruges til yderligere brugertilpasning af robotters adfærd.

Med sit fokus på at øge robotters affektive virkemåde, præsenterer denne synopsis et paradigmeskift i metoden hvorpå vi designer, konstruerer og tester sociale robotter. Metoden argumenterer for at robotter skal kombinere flere enkle kommunikationsmodaliteter til at formidle følelser og lignende komplekse informationer. Dette står i kontrast til at fokusere på en enkelt specialiseret modalitet for at styrke en robots kommunikationsevne. Metoden fokuserer på at kombinere flere forskellige simple kontekstuelle informationskilder for at give et større indblik i omstændighederne for hver interaktion. Dette gøres med henblik på at skabe mere virkningsfulde og troværdige robotter gennem den udvidede forståelse af konteksten. De opsamlede informationer kan bruges af robotterne til at tilpasse sig eventuelle ændringer i den fysiske kontekst og til at imødekomme de adfærdsmæssige krav der ligger i enhver interaktion med mennesker.

På baggrund af vores resultater er det et forslag i denne afhandling, at robotudviklere ændrer deres syn på, hvordan robotter kan anvende kontekstuelle informationer. Dette medfører at kravene sænkes for, hvad robotter skal kunne forstå i den umiddelbare kontekst for sociale menneske-robot interaktioner. En komplet forståelse af den aktuelle kontekst for robotter er muligvis for kompliceret at opnå med den nuværende teknologi. Alligevel giver det mening allerede nu at begynde at bruge de simple kontekstuelle informationer i robotprojekter, da selv helt enkle informationer om konteksten kan anvendes til at forbedre robotters adfærd.

Contents

3 Introduction

3.1 Main concepts

When we humans interact with each other, two things are important to ensure that we are understood. First, we express ourselves using multiple interaction modalities. This means we use combinations of how we gesture, how we move, how we speak, and even how we look, to communicate effectively. Second, we adapt to the contextual demands of each interaction by inferring information from the people we interact with and the places in which we interact [1–3]. By adapting to the contextual demands of our interaction, we try to ensure that we are not misunderstood and that the information we convey is meaningful in the situation. For instance, we may use a soft, low voice to relay sensitive information or a loud, sharp voice when we want to emphasize our frustrations. We also alter our communication strategy if we sense that the information we send out is not being properly received. For example, we may use a comforting smile when the person we are speaking with seems uncomfortable in the interaction, or we may raise a hand to convey the magnitude of what we are saying. Using multiple interaction modalities and adapting to the context are important strategies to use to communicate successfully and creating robots to master these is a crucial step toward improving how they communicate.

We need to change our perspective on how to build robots for human-robot interactions. We should combine multiple expression modalities when designing robots to enable them to communicate more clearly, and we should use multiple simple context measurement methods to heighten robots' level of context adaption. This may pave the way for more dynamic robots and robot behaviors that adapt to the context of the human-robot interaction to ensure an optimal foundation for communication that works across different circumstances.

3.2 A step forward

The affective impact of a robot details how well it can communicate information that may alter a person's current affective state. Such a state can best be described as an experience of a feeling, emotion, or mood at the current moment. A strong affective impact is a desirable skill for a robot to possess, as it indicates that the robot can successfully convey complex information or even emotional states. How the robot is perceived is important, and a strong affective impact adds value to a robot only if there is consistency in how the robot is perceived. For instance, if a robot's behavior evidently excites many people when it is intended to calm people down, its impact is nullified. The examples in this subsection further illustrate how we envision a changed perspective to influence the development and use of social robots.

First, it may be beneficial to pursue context awareness for robots through the behavior-informing features of the immediate environment as even basic context information may be useful for informing affective behaviors. As it is difficult to establish the full extent of people's intentions in an interaction, a beneficial strategy can be to assume that the information will always be incomplete and noisy. However, a robot can still benefit from using this noisy information as a guide in the interaction as long as the information is used only to inform processes at a matching abstraction level. Incomplete information may not provide the robot with a perfect understanding of the situation but may be sufficient to guide the robot in choosing specific behaviors or to adjust simple elements of its current behavior.

Imagine the scenario below.

> *A robot is put to work in a restaurant. The robot is equipped with simple distance sensors and a camera. It detects a person carrying a green jacket. The robot cannot infer the full situation from these details, but it has a simple behavior of looking for humans and jackets. That the robot sees the customer in the open may mean that the customer is looking for the restroom or maybe wants to hang his or her coat. The robot reacts to the customer-room combination and asks the customer if he or she is looking for the restroom.*

The robot may be wrong in a few incidents and right in others, but what the above scenario shows is a robot that reacts to simple contextual cues. For instance, a human holding a restaurant menu is a simple contextual cue that may mean that it would be appropriate for the robot to take the human's order.

Second, we hypothesize that it may be beneficial to let perceived complex emotions in robots emerge from combinations of behaviors and to let meaningful interactions depend less on user states. Simple actions in combination may be perceived as intelligent in the right circumstances. For instance, in Braitenberg's "Vehicles," the simple combination of distance measurements and motor commands may be perceived as curiosity, love, or fright as it enables the robot to approach or disengage from any human who reaches out to it [4]. In contrast, user states can be a weak point in human-robot interactions as each estimated state presents a possibility for the robot to go wrong. This may result in specific behaviors and reactions fitted to specific states being used out of place. Getting the state wrong means that the robot reacts to something that is simply not happening in the interaction, and this may cause the user to disengage from the interaction. Sequential states also indicate that there are specific entry points that demand that the users are at a specific state before the interaction is initiated. State transitions may also require the robot to recognize specific markers to initiate a transition from each state. The attempt to make this detection as robust as possible often results in simplifications of the confirmation cues (e.g., reducing complex questions to "please answer yes or no"), which may again lead the user to withdraw from a deeper narrative.

There may be a simpler way to achieve a meaningful human-robot interaction by combining behaviors triggered by environmental markers. Instead of keeping track of an inner model of the user and the interaction, it may be more viable to use the real world as a model, as Brooks 1991 suggested [5]. Letting the robot focus on contextual markers to trigger simple reactive behaviors may make the users perceive them as acting meaningfully and with complex intentions in the context. The proposed shift in perspective can be summed up in the hypothesis that meaningful interactions and perceived emotions can emerge from combinations of behaviors, and there is no need to use scripted events with inferred user states to develop a deeper narrative.

Imagine the scenario below.

> *The restaurant robot discovers a person holding a menu. The robot has a simple behavior triggered by the menu and the person, and the behavior enables the robot to add a single order. It thus asks the person if there is anything he wants to add to his or her order. Instead of going through*

each step (starters, main course, dessert) while keeping track of the whole process, the robot keeps repeating the simple step. When the person puts down the menu, the robot says "thank you" and drives away.

The above scenario illustrates how a robot can sustain a complex dialogue with a human using a single repeatable behavior. The robot may get the order wrong sometimes but it will refrain from expressing something inappropriate for the current context. The strategy outlines the difference between plans and situated actions as presented in Suchman 87 [6] and as emphasized in that reference, situated actions don't necessarily explude plans. Although it may be preferable to avoid deep dialogue branches, some conversational branches may depend on each other. For example, when a user orders a hot dog, it will make perfect sense to ask him or her a follow-up question about the preferred toppings. In such a case the connecting dialogue states will consist of two manageable steps.

Third, when social robots communicate with humans, utilizing multiple interaction modalities to increase their affective abilities may have advantages. Human-robot interactions can occur across different physical contexts and with multiple human participants. Using only a single interaction modality may prove a perfect fit for one physical context and may be unusable in another physical context. For instance, physical gestures may work well for a robot in close proximity to humans in a small room but may fail to work across a greater distance in a larger physical space. The more an interaction modality of a robot is attuned to the current context the more effective it will be. However, this may come with the inability to work effectively under changed circumstances. Although humans rely more on non-verbal cues when communicating with other humans than when communicating with robots, as stated by Verhagen et al. 2019 [7]. However, using multiple non-verbal communication outlets can also be beneficial for robots. For example, a robot that uses movement and gestures in combination with spoken audio can make it easier for humans interacting with it to perceive it as being hurt rather than having to interpret this from expressive movement alone. Furthermore, using spoken audio alone may not have a sufficient impact on humans to make them believe the narrative of the robot being hurt.

Imagine the scenario below.

The restaurant robot needs to evacuate some customers to avoid a potentially dangerous situation. Its first attempt at getting their attention using its voice in the crowded restaurant does not work. It delivers the evacuation *message again, but this time it uses its body, with shaking movements and heavy gesturing, to catch the attention of the customers and to emphasize the importance of its message.*

The foregoing scenario highlights a situation in which the contextual requirements of the scenario warrant that the robot uses several interaction modalities to successfully convey the gravity of the situation.

It can be argued that the given scenarios simplify very complex human-robot interactions in a real-world example. As such, they are not complete solutions to every aspect of such interactions. Instead, they should be perceived as suggested changes to the control systems of the current social robots and as suggested changes to some of the current research directions. Current research projects involving technical implementations often focus on developing and testing single robot communication features in isolation and often test the robots in strict laboratory conditions while refraining from adhering to any changing contextual requirements.

It may be easy to discover practical disadvantages that will pose problems for the robot in some of the presented examples. The point, however, is that these disadvantages may be countered by adding further loosely coupled single-purpose behaviors. This is opposed to countering any problem by expanding the existing behaviors and thereby potentially creating greater complexity and user state dependence. Our hypothesis is that perceived meaningful intelligent behavior of the robot will emerge from the combination of a sufficient number of simpler behaviors.

3.3 A combination of interaction modalities

Interaction modalities are the different types of communication features a robot can use to express itself as it interacts with humans. In this research, we focused on five different high-level modalities: gesturing, movement and orientation, audio, morphological changes, and anthropomorphic and zoomorphic features. The concept of anthropomorphism refers to the projection of human intentions and motives onto non-human entities such as robots. We humans, tend to anthropomorphize the robots we interact with as found by Fusse et al. 2008 [8]. In this project, the anthropomorphic features pertained to any zoomorphic or anthropomorphic features added to the robot. This may include a head, eyes, tails, and other features that make the robot resemble known animal or humanoid figures in an attempt to invoke anthropomorphism [9]. Humans rely on multiple interaction modalities to communicate when they interact.

Robots may gain potential benefits by similarly utilizing multiple categories of high-level interaction modalities as they interact with humans.

Some people tend to use certain means of communication more than others. Some may use much gesturing while others may accentuate their voice or move their heads energetically to be understood as found by Numan et al. 2019 [10]. Although humans can emphasize certain aspects of their communication methods, they rarely disregard interaction modalities to prioritize a single modality [11–13]. The strategy of not emphasizing a single interaction modality may be beneficial for robots and can make for a less demanding implementation that can apply to more robot projects.

Robots interact with the world through their bodies [14]. Nevertheless, the body and morphology of the robot also influence how it is perceived by the people it interacts with [15]. An anthropomorphic robot may be viewed as having more social presence than a zoomorphic robot as found by Barco et al. 2020 [16]. Robots have bodies, These are not utilities they can switch on and off. The body resides in a physical context and may change that context as it moves around [17]. All these factors influence how we perceive robots, therefore it is logical to assume that multiple aspects rather than a single isolated communication feature of the robot influence every human-robot interaction and must be considered when researching such topic.

3.4 Defining a context

There is a great amount of contextual information that has very little impact on how a robot should accomplish its task. Context information has been defined by Menezes et al. 2014 as information that influences or constrains the way some actions are selected, without being at the center of interest for the task [18]. This means that any robot that aims to use contextual information must filter the information it collects. This is a particularly difficult task for a robot as there is an enormous amount of information that it has to sort through to find the information it needs. Imagine the scenario below.

> *A healthcare robot is being used to vaccinate people against COVID-19. At 1:00 p.m. it enters a large green room to vaccinate a little girl. The girl is having a hard time sitting still and moves anxiously back and forth.*

In this example, the current time, the physical dimensions, and the color of the room have no impact on the vaccination robot's designated task. In contrast, the fact that the girl is moving around is important as it may indicate that the girl is nervous. This information can help the robot change behaviors to calm the girl down or try to make her laugh so she may find it easier to get through being injected. Dennett 1984 defines this as the framing problem of AI and uses an example of a robot that needs to pick up an item in a room containing a bomb timed to go off to highlight how classic AI struggles with deciding which details to focus on and instead tries to incorporate all details. The robot in Dennett's scenario had just finished calculating that removing the wagon would not change the color of the walls in the room when the bomb exploded. [19].

For the human brain, it is a trivial task to filter out all unnecessary information and focus on the vital one. We all receive millions of different inputs each minute, but our brains make sure that only the most important information is considered [20]. Robots currently do not have such a filter. To aid them, we must instead identify the subtle cues in a social context that are usable in relation to what the robot is trying to accomplish, and inform the robot of such. This project investigated the feasibility of increasing context-awareness using input cues from both the physical and social contexts.

3.5 Strengthening context awareness

We, humans are highly skillful in adapting to multiple social contexts based on our perceptions of the people we interact with [21, 22]. For instance, we use a soft voice when we approach someone who is crying or we may behave energetically when someone with a high energy level approaches us. We are even able to adapt socially to other cultures as found in Soltani & Keyvanara 2013 [23]. This is mostly facilitated by an immediate reading of the people we interact with, and should motivate research on the use of immediately available data in an interaction. We also establish recognizable patterns in how other people move, talk, and gesture as we interact with them. We adjust our behaviors so they would match how others behave. Khoramshahi and Billard 2019, and Kühnlenz et al. 2013 found, that robots also have the potential to adapt to how humans behave [24, 25]. Pfeifer et al. 2007 highlight the strength of embodied bio-inspired robots that are robust because of their abilities to adapt [26]. As mentioned earlier, adjusting behaviors on the basis of contextual cues may enable a robot to work across several changed interaction contexts. There are also indications that contextual information may improve the social interaction between humans and robots, as found by Meneze et al. 2014 [18]. This indicates that a robot's affective impact may be influenced by its ability to comprehend contextual information.

Furthermore, Ullrich et al. 2017 found that the task domain influences how a robot's personality is perceived, which indicates that context information can be used as a guide for the adaptation of robot behaviors [27]. The given scenarios show that robots may gain the ability for more personalized interactions and more robust behaviors by adapting to contextual cues in an interaction. However, it may be challenging for robots to determine which attributes are important for their current tasks.

3.6 Embodiment inspirations in our project

When we interact with robots, we tend to anthropomorphize these and project human emotions onto them [8,28]. This anthropomorphism extends to simple moving shapes, as found by Heider and Simmel 1944 [29]. In the Heider and Simmel experiment, the test participants were shown images of two triangles of different sizes moving around. The participants interpreted the scenario as the bigger of the two triangles was a bully who was trying to capture the smaller triangle.

Examples of social robots invoking emotions date back to the 1950s when William Grey Walter created mechanical tortoises, small robots resembling turtles moving around the room [30, 31]. Created to illustrate that simple connected sensor systems could amount to complex intelligent behaviors, they were able to successfully navigate in an indoor arena using simple distance sensors. Although they were not created specifically for social purposes or to invoke anthropomorphism, they possessed some degree of social qualities with their pet-like behavior. Walter's tortoises illustrated how robot behaviors could invoke complex emotions using only simple movements.

Rodney Brooks suggested in Brooks 1989, 1990, and 1991 that using finite symbolic representations of a robot's state and the environment was a limiting bottleneck for the robot. Instead, Brooks suggested that a composition of layered behaviors with a shorter route from the sensors to selected actions could pave the way for a new breed of behavior-based robots that would be able to accomplish more complex tasks compared to the previous generations, including socially interactive tasks [5, 32, 33]. This inspired us to focus in our project on how the immediate reactions of a robot influence how it is perceived. For instance, a robot with a short route from sensing to actions may be perceived as being more responsive and intelligent. Following this strategy, Braitenberg suggested simple robot control systems that portrayed complex human emotions, which again underlines the emergence of intelligent affective behavior [4]. The principle of parallel, loosely coupled processes, as presented by Pfeifer 1996, defines emergent intelligent behavior as an entity facilitated by many simultaneously running subsystems [17]. In our project, we found indications that robot behaviors guided by contextual information, which invoke emotions in those interacting with the robot, may also be successfully facilitated by a combination of parallel subsystems.

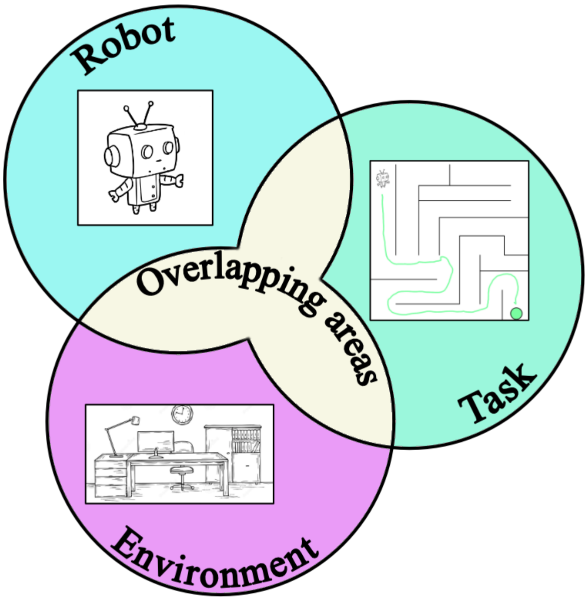

Figure 3.1: The areas of interest influensing affective robot designs that were investigated as a part of this project. These three areas may be vital for emergent behaviors. Such behaviors can be facilitatet as a result of how the robot is constructed morphologically, its control ar- chitecture, and how the environment may be structured. The intersection represent overlap between the different areas.

Both Brooks 1991 and Pfeifer 1996 emphasize the importance of robots having a physical presence (embodiment) and of placing robots in a physical space rather than in a simulated scenario (situated agents), in which both the task and the context influence the perception of intelligent behaviors. Figure 3.1 shows the vital research areas investigated in this dissertation. Each of these areas was investigated to determine its impact on how intelligent affective behaviors can emerge. The high-level areas have similarities with the areas of interest mentioned by Brooks and Pfeifer. Hafner et al. 2002 and Bovet and Pfeifer 2005 state that behaviors enabling obstacle avoidance can emerge as a result of how the robot is constructed morphologically, its control architecture, and how the environment is structured [34,35]. In the same vein, we found that empathy can be evoked or remorse can be communicated by using a combination of embodied behaviors [36]. In Mataric 1991, complex tasks were accomplished with a behavior-based architecture in a robot that could handle the tasks through the synergy between combinations of embodied behaviors [37]. Mataric and Michaud further elaborated on the strengths of combining behaviors to enable robots to handle complex tasks. The mentioned principles build upon the concept in Brooks 1991. He outlines a decentralized model with the strengths of the individual behaviors' ability to store and manage simple representations [5, 38]. Our research had a similar assumption, that it is a good approach to make robots rely on a combination of subsystems providing information through the immediate sampling of the context.

In Breazeal and Scassellati 1999, a context-dependent system was presented to control where a robot focuses its attention [39, 40]. The robot that was used in the experiments was "Kismet," a robotic head that could communicate through multiple interaction modalities and could replicate human emotions. Cynthia Breazeal further embraced the embodiment of an agent as a vital part of human-robot interactions and investigated robot learning inspired by how infants learn. Infants learn through the constraints in the environment and the constraints in what they are presented. "Leonardo," another robot developed at Massachusetts Institute of Technology, was able to infer the internal states of the humans it interacted with and couple these with how such humans felt about an object [41]. It was also able to comprehend social references in the voices of the humans it interacted with, similar to how infants learn through social referencing [42].

This project investigated context awareness through the embodied design perspectives set forth by Brooks and Pfeifer and extended by researchers such as Breazeal and Mataric. We adopted similar perspectives in this project, including the assumption that a finite model of the interaction and context cannot be created. We also emphasized that physical embodiment and situated agent principles are highly influential in invoking emotions in the people with whom a robot interacts. Finally, we found it important to highlight the notion that the perceived robot emotions emerge from and reside in the minds of those interacting with the robot.

3.7 Changing paradigms

This project can be viewed as extending a paradigm shift that started with Brooks and Pfeifer and was followed by Mataric and Breazeal, who extended their research to embodied AI. Through this project, we suggest extending the current paradigm in social-robot development, first on how robots express themselves and how they gain contextual knowledge. This research projects that involve technical implementations are currently dominated by a tradition in which roboticists implement and focus on a sparse set of, often technically complex methods of communicating with humans. This implies that there may be an unused potential in making robots communicate using a wider palette of interaction modalities. Second, we suggest changing the current paradigm that gaining insight into the current context is not feasible as there are too many variables to consider. This often results in contextual information being disregarded in robot projects. We propose that the traditional approach in both these areas be disregarded, and suggest that it may be beneficial to combine multiple subsystems to improve the expression abilities of social robots. By combining simple methods of communication from each high-level category of interaction modalities, the affective impact of the robot may increase.

We also suggest accepting that a single context information system will never be precise. If we tolerate uncertainties in the context classifications made by robots, useful information can be retrieved even from simplistic sensors. The simple sensors informing robots about the context of the interaction through low-resolution information sources are often robust and have high fault tolerance. The provided information does not have to be highly detailed as even simple knowledge about the context may benefit most robots. This project aimed to clarify the validity of this extension to the evolving paradigm shift in social-robot development by focusing on the expression and context awareness abilities of robots.

3.8 Research objectives

The main research objective of this project was to find the answers to the following main question.

- How can we improve the affective impact of robots?

The answer to this question was sought from a hardware and software engineering standpoint and was elaborated on through additional sub-objectives. Figure 3.2 depicts an interaction between a human and a robot. It outlines the parts of human-robot interactions we focused on through our additional research objectives. We investigated the impact of and synergy between the different interaction modalities of robots. This included investigations into the sensing and reaction abilities of social robots and the perceptions of different robot behaviors and engineering aspects. We also investigated the feasibility of using immediate and readily available physical context information, and finally, of measuring and using the immediate cues of the humans involved in the interaction.

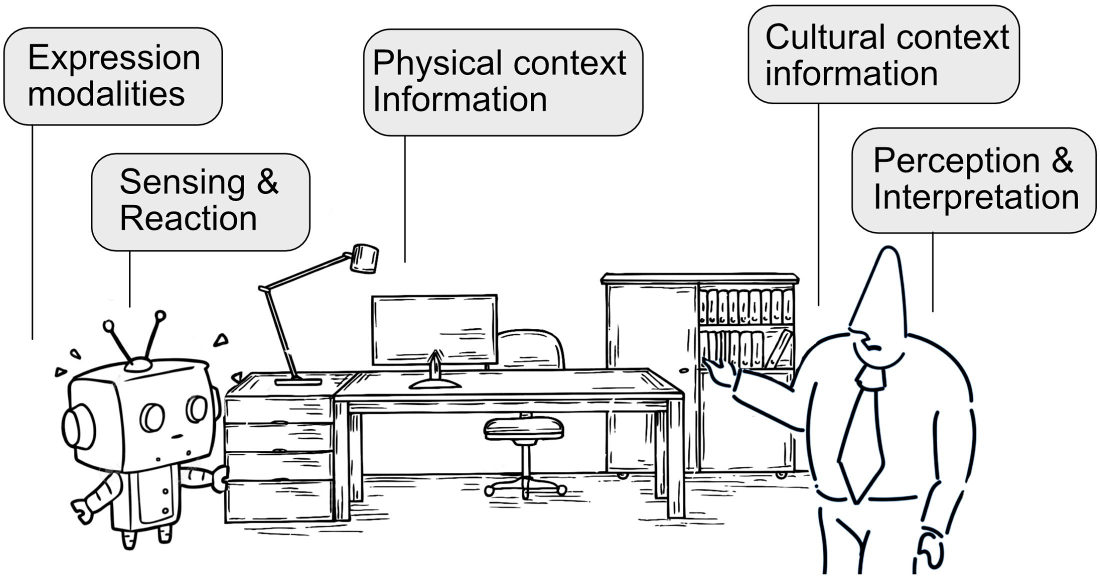

Figure 3.2: The areas of Human-Robot Interaction that were investigated through this project.

The first additional objective was to gain an overview of the affective abilities of social robots. This spawned the research question below.

- How can we define the affective strengths of social robots?

We sought the answer to this question by developing a systematic comparison model applicable to any robot that would provide an objective description of its affective abilities. We did this in our paper entitled "A Systematic Comparison of Affective Robot Expression Modalities," in which we also used the model to gain an overview of the tendencies in the current affective robot research [9]. The resumé of this work is available in Section 5, and the full text is available in Section 11. The paper is included in the proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems.

The second additional research objective was to investigate the affective impact of reactive versus non-reactive behaviors in social robots. This spawned the research question below.

- How does the timing of a robot's behaviors influence its perceived affective impact?

To find the answer to this question, we performed experiments in our paper entitled "On the Causality between Affective Impact and Coordinated Human-Robot Reactions" that determined the effects of events shared by humans and robots during an interaction between them [43]. The results showed a clear impact on how the robots were perceived and that delaying reactions alters their affective impact. The resumé of this work is available in Section 6, and the full text is available in Section 12. This paper is a part of the proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication.

With the third additional research objective, we focused on how robots could measure simple context information and adjust their behaviors so these would match the contextual demands of an interaction. This spawned the research question below.

- How can social robots be enabled to retrieve and use contextual information to improve their affective impact?

We sought the answer to this research question by investigating the feasibility of using low-detail context information to drive behavior selection in social robots. In our paper entitled "Adaptable Context-Based Behavior Selection in Autonomous Robots" we presented a context representation system for a social robot. The system could facilitate behavior prioritization to infer the best behavior for previously unvisited physical contexts [44]. The resumé of this work is available in Section 7, and the full text is available in Section 13. This paper is currently under review for publication in the 2021 IEEE/RSJ International Conference On Intelligent Robots and Systems.

We also helped find the answer to the third additional research question, through our paper entitled "A Minimalistic Approach to the User-Group Adaptation of Robot Behaviors using Movement and Speech Analysis" [45]. In this project, the main focus for contextual awareness was enabling robots to use the human cues available within the initial minute of an interaction. We investigated the minimal number of speech and movement characteristics needed for a robot to distinguish its users, and how these numbers are influenced by human-robot interactions across multiple contexts. The results showed that robots could use these types of data to distinguish individual humans in the context, but that the information was insufficient for robots to do the same across multiple contexts. The resumé of this work is available in Section 8, and the full text is available in Section 14. This paper is a part of the proceedings of the 2021 30th IEEE International Conference on Robot and Human Interactive Communication.

Although the works mentioned below are not included in the main contents of this dissertation, we may also briefly touch on them.

- "Robots Can Defuse High-Intensity Conflict Situations," included in the proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems [36].
- "Augmenting the Audio-Based Expression Modality of a Non-Affective Robot," a part of the proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction [46].

3.9 Content overview

The following section (Section 4) presents an overview of the related previous projects and the suggested design principles for affective robot development. The subsequent Section 5 summarizes the findings of the first project reported in this dissertation regarding the impact of using multiple interaction modalities in human-robot interactions. Section 6 details the findings of the second project regarding the coordination of a robot's reactions to outside stimuli. Section 7 outlines the findings of the third project on the feasibility of using immediate context information to drive behavior selection. Section 8 presents the findings of the final included project on the use of contextual cues from the humans interacting with robots. Section 9 discusses each design principle with a focus on the engineering aspects of each principle and hypothetical evaluations of each principle. Section 10 presents the conclusion arrived at on the basis of the obtained project results. Sections 11, 12, 13, and Section 14 present the full text of each of the included papers. Finally, section 14 contains the numbered references of this dissertation.

4 Principles of affective robotics

This section reviews the important milestones in the relevant literature starting with a short chronological review of affective robotics research. The section then presents references to state-of-the-art research projects and highlights the results obtained on each of the high-level topics investigated in this project. With each topic, we highlight how our research relates to the presented approaches or how we complement the previous findings. Some references used in this section may appear in the papers included in this dissertation alongside other new relevant research references that have emerged since the original papers were published. The included references cover each high-level topic investigated in the project and we discuss how the findings were distilled into the suggested design principles for affective robots that were used as guidelines throughout the project.

4.1 Emphasizing anthropomorphic interpretations

In this subsection, we focus on how anthropomorphism was used in previous robotics projects. This includes the tortoise robots created by William Grey Walter in the 1950s as an early example of social robots [31, 47]. They had some social skills through how people interpreted their behavior, but examples of affective robotics and of human-robot interaction date even further back. "The Turk," a seemingly autonomous chess table developed by Baron Wolfgang von Kempelen in 1769, was introduced as the Chess Playing Automaton [48]. The machine was touted as capable of besting all its challengers. In reality, the contraption's inner machinery consisted of a hidden human controlling the chess moves. The interaction between the players and the contraption can be seen as an early version of "Wizard of Oz" interaction studies (named after an animatronic figure in the movie with the same title) [49]. The word "robot" was used for the first time in a 1920 play by Karel Chapek about a constructed mechanical humanoid. Through the events of the play, the robot is shown to possess human emotions and eventually revolts against its creator [50]. "Shakey" the robot developed within the period from 1968 to 1972, was the first mobile robot that used a layered control architecture to navigate through a simple scenario. The robot was an early example of the application of an autonomous artificial intelligence (AI) that was able to assess the outcomes of its actions [51].

Brooks 1986 presented a new approach to AI, abandoning the idea of creating a symbolic representation of the world and instead began using the world as a model. As he stated, *"The world is its own best model"* [52]. This drastic departure from mainstream AI also included a departure from relying on calculated actions as used in a symbolic representation of the environment. Instead, the robot would use layers of behaviors with actions and measurable reactions in its real-world environment. In 1999, Sony Electronics Incorporated introduced AIBO, a dog robot developed solely to function as a consumer version of a social robot. AIBO's controller was also behavior-based, but the number of behaviors was extended to hundreds of simultaneously running parallel behaviors [53]. Pepper, a robot created by SoftBank Robotics, is frequently used in research and is an example of a robot being used to provide social functionality outside academic projects [54]. Pepper is often used in the industry to welcome guests to new venues and to offer information about conferences, etc. With its humanoid form, Pepper is also used for various research purposes, such as in therapeutic settings [55]. The Nao Robot was also developed by SoftBank Robotics. The humanoid robot has a toy-like appearance but can be programmed to replicate complex human behaviors [56]. It has been used in research projects on a wide array of topics, such as affective touch in Andreasson et al. 2018, the perception of mechanical sounds in robot gestures in Frid et al. 2018, and improving the balancing skills of robots with reinforcement learning in Tutsoy et al. 2017 [57–59]. "Paro the Seal" is a successful non-humanoid social robot developed in 1996. It facilitates therapy and care with an emphasis on helping the elderly using visual, audio, and tactile sensors [60, 61]. Hanson Robotics developed its humanoid robot "Sophia" in 2016 as a robot that resembles a human to the extent that the limitations in technology then would allow [62]. The robot is featured in both academic and cultural circles. It is being used to further both AI and human-robot interaction research and engineering research on topics such as arm motion generation in Park et al. 2018 [63]. The robot was a keynote speaker at the 28th IEEE International Conference on Robot and Human Interactive Communication [64].

Complementing these findings

Some of these examples of social robots are the results of long-running and well-funded projects. Such projects may have had a tendency to design robots inspired by human physicality. The projects replicated the human physical properties and human-like behaviors in robots to strengthen how they interact with humans. The hypothesis behind this approach is that humanoid robots and their human-like features may better enable humans interacting with them to form anthropomorphic interpretations of the robot's actions. That is, it is believed that the affective impact (the perceived strength of the invoked emotions) will increase as the robot approaches human likeness. However, recent experiments using zoomorphic robots have shown that non-humanoid robots may equally invoke emotions. This was shown in Pütten et al. 2014, in which the participants responded equally to a video portraying the mistreatment of both small dinosaur robots and humans. The findings were verified by measuring similar neural activation patterns in the brains of the participants as they watched the video [65]. The findings shown in these historic examples made us come up with the hypothesis below.

- Anthropomorphism may be emphasized to increase the affective impact for robots.

It should be stated that by emphasizing anthropomorphism we do not mean that a robot must resemble or use human-inspired features to heighten its communication abilities, in fact, the aim of this project is to advocate for the exact opposite. The effects of anthropomorphism can be experienced even with the movements of simplistic triangles, as the Heidler and Simmel experiment showed [29]. This means that roboticists do not have to use humanoids or human-inspired features to achieve their desired effect, and that is a good reason for investigating how anthropomorphism can be used to improve the affective strengths of non-humanoids using simple behaviors and technical implementations.

The results revealed in the included examples were groundbreaking but may be difficult to directly apply in some of the current robot projects. We investigated how simple robots that are not using multiple interaction modalities can benefit from augmenting the areas that are not utilized in their current setup. With the augmentation of non-utilized areas of robot communication modalities, as we discussed in our 2019 paper, we also focused on the immediate applicability of our results [46]. These results may be directly applied to the current breed of robots that were not initially designed for social interaction. In this project, we also created many smaller prototypes that included as many different interaction modalities as possible. As stated by Zamfirescu-Pereira et al. 2021, it may also be beneficial to follow more design-based research patterns when building robots, as insight can be gained by making many prototypes [66].

It is also evident in the references, such as the William G. Walter tortoises, that even simple behaviors can make robots appear intelligent and can enable them to have a strong affective impact on the humans who perceive them. In our project, we created prototype robots under the assumption below.

- Perceived complex emotions can emerge from the combination of simplistic behaviors.

A good example of the above assumption is given in Erel et al. 2021. These researchers investigated how the simple behaviors of two robots playing ball with a human participant can influence how the situation is perceived by the participant [67]. They found that making the robots throw the ball at certain angles toward the participants could make the participants feel rejected, ignored, and meaningless. This effect, that simple behaviors in certain contexts can invoke a strong affective impact, has been used in all our experiments. Affective communication can have a large impact on our society, as Bana et al. 2021 and Tsai et al. 2021 demonstrate with a robot that uses different behaviors to encourage humans to sanitize their hands [68, 69].

4.2 Utilizing multiple interaction modalities

In this subsection, we discuss how we focused on different categories of interaction modalities. We give examples from each category and explain how they influenced us to focus on the synergy among the categories. The expression capabilities of social robots can be roughly divided into the following five high-level interaction modalities:

- The robot's motion and orientation.
- The robot's general morphology.
- The robot's posture, gaze, and gestures.
- The robot's audio-based communication abilities.
- The robot's anthropomorphism-invoking features.

This subsection includes references from research projects with robots that emphasized one or more specific modalities.

The motion abilities of a robot influence how it travels from point A to point B and the movements that it performs on the spot. This is often referred to as the Kinesics. The orientation abilities reflect how well the robot directs its attention toward either a person or an object in the environment. Simmons et al. used Laban movement analysis to create expressive motion in a dancing robot [70]. Rudolf Laban (1879-1958) was a choreographer who created a notation system for expressive movements, and the system has been complemented by Perez and Barakova 2020 as the foundation for designing expressive interactions for embodied objects [71]. The same system has also been used by Sharma et al. 2013 and Bevins et al. 2021 to create affective locomotion for flying robots and has been used by Knight et al. 2014 to design expressive motions for mobile robots [72–74]. Expressive movement may even be effective for priming humans to behave differently in immediate subsequent scenarios [75]. Mead and Mataric 2013 trained a neural network to identify the relationship between proxemics and user-initiated social actions [76]. Kitagawa et al. 2021 designed the movement system of an omnidirectional robot inspired by how humans move toward a task-specific target. They tested their movement system on 300 observers and found that people interpreted the robot's movements as a natural fit for the robot even though the robot rotated freely while traveling straight [77]. The specifics of movement and orientation also include the acceleration changes of the robot as outlined in Saerbeck and Bartneck 2010 [78], the changing travel speeds of the robot as described in Yoshioka et al. 2015 and Knight et al. 2015 [79, 80], the variation in the directional changes as used in Fernandez and Bonarini 2017 as a part of a system developed to enrich robot movement for conveying affective information [81], and the specific orientation [82]. In Bethel et al. 2009 they enabled the robots to express attentiveness by using slow movements while sustaining the orientation toward the test participant [82].

As stated in our 2019 paper [9], the morphology of a given robot describes its physical appearance. The previous projects on the impact of different morphologies include that by O'Brien et al. 2021, who investigated the impact of using different morphologies for therapeutic robots for children. They found that the kids preferred to interact with a pillow-type robot by stroking, holding, and cuddling it [83]. In the 2016 project by Cha and Mataric, lights were used in combination with audio for a robot to signal for help in a cooperative task with a human [84]. Miller et al. 2015 aimed at facilitating a close natural interaction between a human and a pet-like robot. They developed a tactile sensor that looked like animal fur to encourage tactile interaction [85]. In Schellin et al. 2020, fur was also the material that was added to an AIBO robot. They found that adding fur made their participants dislike the robot when it was framed as a puppy. However, fur did make the participants perceive the robot as being less scary [86,87]. The morphology also entails focusing on the build material of the robot, as evident in the materials used by Stiehl 2006 and Sefidgar et al. 2016 [88,89], and the physical proportions and shape such as those utilized by Boccanfuso2015 et al. 2015 and Singh et al. 2013 [90, 91].

The posture, gaze, and gestures category describes how a robot communicates using onboard movement or through gaze directions. Block et al. 2021 constructed a hugging robot that adjusted its embrace and posture to the position and size of the test participant [92]. Knight et al. 2017 used a robot chair to communicate when a human could pass it and found that motion alone could successfully convey the intents of the robot [93]. In the 2020 paper by Panteris et al., the focus was to generate waving patterns for non-verbal communication [94]. Their presented system could successfully generate waving movements based on different expression characteristics. Each of the patterns they investigated could be sufficiently executed by a six-degree-of-freedom robotic arm. Projects investigating affective gesturing use a variety of features to convey emotions, including body postures, as used in Cohen et al. 2011 [95]. They also include gesturing as the dog tail presented in Singh 2013, or the human-inspired arm gestures developed for a robot storyteller in Xu et al. 2015 [91,96]. Knight et al. 2012 used affective gesturing to enhance the expressions of a robot actor [97]. Rincon et al. 2018 mapped values from Mehribian and Russells' Pleasure-Arousal-Dominance (PAD) emotional model onto expressive gestures for a robot arm [98, 99]. Lee et al. 2013 used different postures for a robotic room divider to convey whether or not the test participants should approach it [100]. Yu and Tapus 2020 used a generative adversarial network to dynamically create movements using audio from speech as the input [101].

The audio-based expression abilities of a robot encompass the audio originating from the robot. This includes both the sounds made by the robot by physically interacting with the environment and any artificially generated audio used for communication purposes. Robinson et al. 2021 designed artificial movement sound and used them to alter how a robot was perceived [102]. Rossi et al. 2020 used audio-based communication features in combination with movements to gain the attention of children who were getting vaccinated. They found that their robot could lessen the pain of the vaccination shot [103]. In Pipitone et al. 2021 the voice of a robot was used to express the reasoning behind a robot's decision process [104]. The transparency of such a setup provided anyone a better insight into the current state of the robot and could potentially create a stronger bond between it and anyone who interacts with it. Winkle and Bremner 2017 used a robot's voice to convey emotions [105]. Becker-Asano et al. 2009 tried to create the best synthetic laughter to improve social interaction with a robot [106]. Lui, Samani, and Tien 2017 used soundscapes to brighten the mood of the participants who interacted with their robot [107].

The anthropomorphism-invoking features of a robot are the features that are added to make the robot appear like a recognizable figure, with the aim of projecting human emotions onto the robot. We tend to anthropomorphize both robots and virtual agents as stated by Darling et al. 2015 and Zawieska et al. 2012, and according to Natarajan and Gombolay 2020 the level of anthropomorphism projected onto a robot influence how much we trust it [108–110]. Hover et al. 2021 found that their research project participants' attitudes toward different robots changed as they varied in Gender and human likeness [111]. Marchesi et al. 2021 found that the tendency to anthropomorphize non-human agents influenced how quickly humans accepted mentalistic descriptions of a robot [112]. Examples of human-inspired robots include the previously mentioned Sophia by Hansen Robotics, Pepper and Nao by SoftBank Robotics [54,56,62]. Becker-Asano and Ishiguru 2017 also used the humanoid robot "Geminoid F" to investigate expressive facial expressions [113]. "Barthoc Jr." is another example of a humanoid robot. In the 2021 project by Faraj et al., the robot consisted solely of a human head replica that used its 25 facial muscles to emulate human expressions [114]. Hegel et al. 2011 investigated meaningful cues and signals for robots in an interaction and stated that the appearance of a robot could also include intended cues [115]. Collins and Mitchinson 2015 investigated the impact of familiarity in body language and other emotional expressions from zoomorphic figures [116], and Breazeal et al. 2004 used "Leonardo," a small teddy bear robot, to investigate human-robot collaboration [117]. The previously mentioned Aibo and Paro are both examples of zoomorphic or animal-inspired robots that are used for recreational and therapeutic purposes [61,87]. Also, In Canamero et al. 2016 the robot controller was designed with a focus on how the robot was perceived in interactions. They augmented a humanoid Nao robot to successfully provide diabetes management to children through different social interaction styles [118].

Complementing these findings

Similar to what is stated in our 2019 paper, many project robots were created and tested with a focus on few interaction modalities in their communication abilities [9]. This means that the tests performed on each of them could have emphasized one feature and tested it alone, in isolation from other features. This approach makes sense as it allows the researchers to obtain a clear result from the experiments. This strategy was also inherited from centuries-old research traditions embracing isolation of the investigated subject and elimination of any variable that could influence or diffuse the result. We align with such an approach to testing and verification in most of our papers that were included in this dissertation. However, this project also suggested the change shown below.

- Multiple interaction modalities should be utilized when designing affective social robots.

It may very well be the case that some of the other factors influencing human-robot interactions are infeasible to exclude in the first place. The potential tendency in the current research strategies to focus on a few interaction modalities in social-robot communication in isolation may be problematic. When a robot interacts with humans, the humans may not be able to disregard other undeveloped features of the robot. For instance, it may be difficult for a human to evaluate the gestures of a robot consisting only of a torso, with wires hanging out everywhere, or it may be impossible for a human to decide between different suitable voices for a robot based on a presented static image or a virtual character. For the robots that we designed for experiments in human-robot interaction, we designed them with abilities covering multiple interaction modalities.

4.3 Coordinating a robot's actions

This subsection highlights examples of how the temporal aspects of different robot behaviors impact how robots are perceived in an interaction.

Mirnig et al. 2015 analyzed the data obtained from 201 human-robot interaction experiments to see how quickly and in what manner the humans reacted to the robots when the latter made mistakes in the experiments. They found that the humans reacted with social signals with a delay of 1.63 seconds to a robot that performed an erroneous action [119]. They found that there was a difference in how much the humans reacted when the robot crossed the border of an implicit social norm and when the robot malfunctioned due to a technical error. Using pauses to convey a status for a robot in an interaction was also included in the 2014 project of Bohus and Horvitz [120]. They used a Nao humanoid robot in a physical interaction in an office space and attempted to control and predict the disengagement of the interacting users. They found that introducing filled ("uhm" sounds) and non-filled conversational delays can be successfully used by the robot to convey uncertainty and to keep the user engaged. Langer et al. 2020 highlighted how the movements of robots can prime how humans move (that we change movement speeds when interacting with a slow-moving robot), and indicated that the different timings of robots influence user satisfaction [75].

The timing aspects span many facets of how robots behave in human-robot interactions. These include robot movements, the general timing of how robots speak, and also the timing of robot errors in an interaction. Huber et al. 2008 focused on the timing aspects of a human-robot collaborative task. They used a situation in which a robot would hand over objects to humans and changed the timings between the experiments to isolate the impacts of the specific timings. They found that to successfully succeed in object handovers, both the human and the robot in the interaction must agree on a shared timing for all the movements involved in the task. They likewise found that increasing the stability of the robot's movement made the participants feel safer in the interaction [121]. When humans and robots need to agree on a shared timing in a cooperative task, it often requires the humans in the interaction to adapt to the robot's movements. Vannucci et al. 2019 hypothesized that humans from different cultures would differ in their level of willingness to adapt, but after conducting studies in Italy and Japan, they found no evidence of such a difference. This may indicate that it may be feasible to determine a proper movement control scheme for robots that will be applicable across cultures [122]. Macarthur et al. 2017 investigated how the speed and proximity of a robot influence how much a human trusts it (measured using the Human-Robot Trust Scale) [123, 124]. They found that the physical presence of a robot, how fast it moves, and how close it is to the human change the human's level of trust in the robot.

Sharing an experience with a robot may also impact how a human feels about the robot. Bing and Michael 2012 investigated how humans who shared a stressful experience with a humanoid robot, felt about the robot. They found that sharing the experience could aid humans in disregarding the undesired effects of the uncanny valley. The results of the experiments conducted in the project showed that the test participants were more likely to prefer humanoids they had previously shared a stressful experience with to robots they had shared a non-stressful experience with [125, 126]. Deshmukh et al. 2018 investigated how the speed and amplitude of a robot's gestures influence how the robot is perceived by the humans who interacted with it [127]. The experiments used a Pepper robot and changed various aspects of its gestures while groups of humans rated the robot's behavior in terms of anthropomorphism, likability, animacy and intelligence, and Safety [54, 128]. The ratings showed a connection with the speed and amplitude of the gestures. I was also found that there was a connection between different types of gestures and specific scores. For instance, welcoming gestures with open arms resulted in higher scores for anthropomorphism.

Complementing these findings

This project investigated how the coordination between the reactions of a human and a robot as they experience the same event influences how the human perceives the robot. The references show how various aspects of robot behaviors, such as the speed, proximity, and timing of a robot's actions (often referred to as the proxemics and Chronemics) influence how the robot is perceived in terms of trust, likability, and safety. Most of the interactions in the projects reported herein used a humanoid robot in a cooperative setting in which a human and the robot needed to achieve a common goal. To widen the general knowledge of these findings, this project focused on whether or not the same results could be gained from using non-humanoids in a non-cooperative setting.

Each action that a robot performs following an event also provides opportunities for the robot to establish a connection between the action and the event. As robots can react much faster than humans, this project investigated the role of the exact milliseconds between the event and the reaction. In our 2019 project, we investigated the perceived impact of coordination between humans and robots reacting to the same event, and the impact of delaying reactions within the immediate window of milliseconds after the occurrence of a shared event [43]. Through our experiments, we found indications that support the suggested design principle below.

- Interacting is reacting; the temporal aspects of an interaction can alter the affective impact of robots.

There are multiple indications of a connection between the impact of a robot's behavior and the timing of its initiation in an interaction. Being able to utilize different timings for various robot tasks may have an impact on the field of social robotics. For instance, it may be beneficial for a rescue robot to move slowly when trying to comfort someone in a rescue scenario while a robot may be more effective in conveying danger and warning bystanders by reacting and moving swiftly. In our 2019 project, we also investigate the different interaction modalities when used in a reaction in a high-intensity conflict scenario [36]. The results indicated that the important aspect of conveying affective information in the interaction is that the robot reacts, not necessarily how it reacts or what communication outlet it uses to react. This was also shown in our 2019 project, in which reaction to outer stimuli was a vital part of the introduced model of affective expression modalities (MOAM) [9].

4.4 Controlling the perception of robots

The project was influenced by research in the field of cognitive psychology. This subsection discusses how the current findings in psychology have influenced what we include and exclude when evaluating experiments in robotics. The research and experiments conducted in human-robot interactions are often dependent on humans evaluating their affective state (how they feel) when interacting with robots in different scenarios. The participants who evaluate their own feelings may introduce errors in the results of these experiments. Self-reporting bias is known to happen in every branch of research from organizational research to social science research [129, 130]. Problems including social desirability are evident in both online and offline surveys [131]. Some branches of neuroscience and psychology research have attempted to clarify how we humans classify emotions and evaluate our own affective states. Quigley et al. 2021 found that it is difficult to measure interoceptive signaling in humans [132]. This indicates the difficulty of verifying internal human processes because we have limited tools to do this. It is also complex because most of these processes are initiated subconsciously, without us realizing it. The process that influences how we perceive body signals relating to our state of wellbeing is called interoception as stated in Kleckner et al. 2017 and Craig et al. 2002 [133, 134]. Hoemann, Devlin, and Barrett 2020 state that emotions in infants materialize as abstract conceptual categories and are classified and explained using language which is gained by the infants as they develop. This indicates that the large plethora of emotions felt by infants may start fine-grained but are evidentially classified as the basic emotions (happiness, sadness, disgust, fear, surprise, and anger according to Eckman 1984-1992) because of how they are explained in our language [135, 136]. This may increase the difficulty of using language to verify the affective impact of robot behaviors in human-robot interaction experiments. Barret 2017 revised the previous basic emotions explanations and presented theories on emotions as constructed. The concept presented here is that emotions are constructed on the basis of previous experiences and as a solution for the interoception process to facilitate energy preservation in our bodies [137]. Our bodies and minds construct emotions to best cope with our current context and to ensure that our body will preserve the greatest amount of energy possible for our handling of our expected future experiences.

Complementing these findings

The findings in these references are state-of-the-art neuroscientific theories that may or may not become the mainstream explanations of how emotions and affect function. Although these are proposed theories and not necessarily facts, the changed perspective on emotion research and the problems found in self-reporting techniques inspired us to focus on the following problematic issues in our approach to experiments in human-robot interaction.

- When and how humans perceive robots and evaluate robot behaviors are factors that are difficult to control.

The aforementioned theories may have an impact on the methods we use to obtain results from human-robot-interaction experiments. These include both what we choose to measure and how we measure it in experiments. It may also impact the amount of relevant contextual information we decide to include in our evaluation of studies.

An important aspect of the theories on constructed emotions as a part of the interoception process is that this process is a subconscious one, meaning you cannot turn it off. Our brain classifies the current context and situation to best handle energy preservation for our body with every included contextual input. This includes both conscious and subconscious inputs. We become aware of an emotion when we sense it as an affect. We then experience this affective state as either pleasant or unpleasant and with a high or low arousal level, as pointed out by Mehrabian and Russell 1974 and extended by Mehrabian 1980 [99,138]. This also means that when researchers use human participants to evaluate robot features, this evaluation may begin sooner than the actual experiment.

The theory of the inclusion of subconscious stimuli in the construction of our immediate emotions may have an impact on how researchers should perform experiments. When researchers ask the participants to focus on a single communicative feature of a robot, it may be difficult to isolate the effect of the feature in the result. This theory indicates that there may be occluded inputs in the interaction between a human and a robot, which influences the constructed emotion. For instance, it may be difficult for the test participants to rate a robot's human-inspired arm gestures if the sounds of servo motors and turning cogs influence the interaction.

4.5 Gathering low-resolution behavior information

This subsection presents our inspiration for focusing on adapting the behaviors of robots to different users by using the attributes obtained in the immediate context and within a limited-duration interaction. As we pointed out in our 2021 paper, different personality factors have already been used in different robot research projects to adapt a robot's behavior to different user segments [45]. Such projects include that by Momen et al. 2018, who presented a robot that expressed extroversion using gaze movements [139]. Matching the levels of extroversion of the participants (measured pre-experiment) with that of the robot had a positive effect on how the participants perceived the robot. This was also found in Craenen et al. 2018, in which the test participants preferred robots that possessed traits similar to their own [140]. Ondras et al. 2020 used audio as an input to generate upper-body movements on a Pepper robot [54, 141]. This is an example of easily attainable basic contextual information used to adapt robot behaviors in the immediate interaction. Hiolle et al. 2014 used a syntehtic emotion model the reacted to the modelled arousal level. The arousal level was influenced by the amount of human contact. Specific levels would make the robot detect a learning challenge and as a reaction it would initate human contact [142]. Menezes et al. 2014 discussed how to integrate context-awareness in social robots. They suggested a system that runs in the background and detects context changes. The changes trigger the activation of new behaviors in the robot's main execution loop [18]. Syrdal et al. 2006 noted in their experiments that the level of extraversion of the participants related to how much they would tolerate from a robot that made their interaction uncomfortable [143]. Xu et al. 2012 investigated the role of the social context in the interaction and found that how well the robots were accepted by the project participants depended on the specific social context [144]. Many of these references used questionnaires to discover personality traits. These often included the "Personality Inventory" established by Eysenck 1965, a series of questions that measures the level of extraversion (among other personality factors) found in the participants [145, 146]. Such a questionnaire was also used in the 2008 project of Tapus and Mataric, who investigated the behavior adaptation of a rehabilitation robot following specific personality traits [147]. In Tanevska et al. 2020 they also adapted the behaviors of a robot to maximize the pleasantness of the interaction for its peers. The adaptation was also guided by synthetic emotion algorithms that controlled the needed level of response and input [148]. The immediate information that is only obtainable in the interaction can also be viewed as a reference to the concept of situated knowledge in Harraway 1988 [149].

Syrdal et al. 2009 presented the "Negative Attitude towards Robot Scale" for use in discovering biases toward specific robot behaviors [150]. The researchers did not find a correlation between the evaluation of the behavior and the scale. However, they did discover a correlation between the scale and the comfort level of the participants when interacting with the robot. Yamada et al. 2021 looked at the scenario that could lead to children abusing robots and identified five different subsequent factors that could escalate a negative scenario [151].

Batrinca et al. 2012 analyzed videos of test participants in collaborative settings and found that their approach could automatically detect the extraversion level of the participants [152]. The level of extraversion of the human participants was also determined by Pianesi et al. 2008 who used video analysis of human-to-human interactions in their project [153].

Building on these findings

The references show that progress has been made in adapting robot behaviors to the personalities of the users by focusing on various types of contextual information. There is a tendency in some of the current social-robot projects to focus on the major personality factors (e.g. the "Big Five") and to use these to adjust a robot's behavior. This approach is feasible as the personality factors are well-researched cognitive-psychology concepts that are measurable through experiments. However, for robots to gain insight into any of the personality factors of the humans they interact with, a thorough data-gathering process is often needed. This process usually includes a series of participant questions asked prior to the experiment, which can impact the immediate applicability of the results. This inspired us to investigate the principle below.

- Combining multiple sources of low-complexity context information can increase the contextual insight.

It is challenging to gain an overview of complex personality factors solely from an interaction between a human participant and a robot. This explains why most research projects focusing on personality factors attempt to gain insight into such factors ahead of the interaction. This is often accomplished using questionnaires following the methods described in previous psychology studies. Examples of such are the questionnaire on personality traits used by Cattell 1943, the Personality Inventory presented in Eysecks 1965, Eysenck 1975, and the Big Five Factor personality model used by McCrae and John 1992 [154–157]. Although these approaches can provide the necessary insight into the personalities of the participants, using them can make it more cumbersome to directly apply the research results in the current social robots. For instance, robots in real life usually do not ask humans to fill out a questionnaire before they interact with them. This project focused on using the immediate but less detailed contextual information available in the interaction. This entailed investigating the feasibility of using contextual information to reinforce behavior adaptation across different contexts. Using the personality information available in the immediate situation is roughly what humans do. We often fit aspects of our behavior to the characters we interact with by reading the immediate situation. There is an opportunity for social robots to benefit from using a similar strategy.

4.6 Using physical context information

This subsection highlights examples of how the information gained from the physical work environment of robots has been used to gain better context understanding. Utilizing knowledge on the physical context of robots may improve human-robot-interaction scenarios, as shown by Roger and Christensen 2012, who presented a system that helped robots understand the semantic meanings of the physical placements of different objects [158]. Torre et al. 2020 investigated the relationship among the task, work context, and voice of a robot, and found that the robot's designated task is highly influential in what voice type the participants matched to a specific robot [159]. Coupete et al. 2016 tried to make a robotic arm move in an acceptable collaborative manner that aligns with the requirements in the environment and with the users' expectations [160]. Aliasghari et al. 2021 investigated how arm-motion kinetics and gaze types influence how a humanoid robot is perceived and found that arm movements can affect the robot's perceived confidence and eagerness to learn. They also found that the type of gaze affects the robot's perceived level of attention to its tasks [161].

Cosgun and Christensen 2018 added context awareness to enhance a person-following robot. The robot predicted targets of future human interaction by looking at the velocity of the human it was following [162]. Using movement and audio to sample details about the users has also been the focus of previous projects. Nakadai et al. 2003 used both to determine the active speaker in a multi-speaker environment. The performance of their developed system depended on the accuracy of localization. They found that motions directed toward the sound source improve the recognition of the active speaker as it strengthens the system's ability to separate the speaker from other audio sources [163]. Zafar et al. 2018 used machine learning to classify the movement patterns captured by an RGB-D sensor. They also classified the levels of extroversion, agreeableness, and neuroticism traits [164]. They used a humanoid robot in different role-play scenarios and reached high accuracy in the automatic assessment of personality traits. Lera et al. 2017 used the acoustic properties of an environment to facilitate differentiation between indoor contexts [165]. The project trained neuronal networks on ambient audio signals to classify the contexts.

Improving the communication skills of a robot was the focus of Xiao et al. 2016 [166]. They tried to enhance human-robot interactions by including communication through natural body language. They created a robot that understood the meaning of human upper-body gestures. The robot communicated with facial expressions, movements, and spoken language. Liu, Wang, and Wang 2018 focused on pose estimation to facilitate the taking of context-aware safety measures for users working near an assembly robot. The system used multiple camera sensors of different types to record and classify assembly poses. The results were used to determine the intentions of the users in the vicinity of the robot [167].

Building on these findings

The references indicate a general momentum toward using physical context information in social robotics. This is evident across a diverse set of robotics research projects. These projects utilize different approaches to retrieving context information, and the detail level of the retrieved information varies with each approach. Although retrieving physical context information is attainable, raising the detail level of the gathered data is often achieved only with an elaborate sensor setup. Furthermore, a higher detail level also increases the complexity of the needed data-processing algorithms. Inspired by these findings, this project investigated the veracity of the statement below.

- Basic context information may be useful for informing affective behaviors.

In real life, we humans cannot switch off our input-sensing capabilities. We sense and experience all the time while our brain filters out irrelevant inputs [20]. This inspired the idea of a new breed of context-aware robots. Similar to the interoception process, these robots will continuously evaluate their context using easily attainable sensor data. They will benefit from not demanding perfect information about the context. Instead, they will accept and use imperfect context knowledge to guide the selection of robot behaviors based on simple sensor information. The simplicity of the sensor setup will allow such a process to run at a high refresh rate.

There may be an unused opportunity in focusing on using the sensors present in most social robots to gain a rough estimate of the physical context. This low-granularity context information can be used to guide the selection of behaviors to align with the current context. In our 2021 paper, we focused on the physical dimensions of the context environment and adapted a robot's behavior to the preferences of the test participants in the different physical contexts [44].

The main strategy entails accepting that getting a complete context overview may be infeasible. Instead, using simple sensors and basic context information may be usable in guiding behavior adaptation. Furthermore, as such information may be easily fetched at a high refresh rate with simple sensors and low processing requirements, it is easy to apply in most robot projects. Although these data may often be inaccurate, the high refresh rate makes up for various error readings over an averaged result. A simple setup also presents the opportunity to add further sensors and combine multiple sources of information. When accepting inaccuracies from single simplistic sensors, such a setup may be beneficial in raising the detail level of the gathered physical context knowledge.

4.7 Summary

The research papers presented in this section discussed the progress made in the main focus areas of this project. The first part included historic examples that inspired this project to focus on emphasizing anthropomorphism and the perceived emotions that emerge from using a combination of different behaviors. The second part presented examples of the impact of using different interaction modalities. It highlighted a tendency in the current research to focus on few interaction modalities in human-robot communication. This inspired us to focus on how human-robot communication influences an interaction when robots make use of multiple modalities, and on including more factors of the interaction besides the communication features. The third part presented references that utilized the coordination of behaviors and highlighted how we used it in this project. The fourth section referenced various research results from state-of-the-art neuroscientific research, which included theories on emotions as constructed entities. This had an impact on the data-gathering processes in this project. The theories may or may not be proven correct, but it may still be infeasible to ask the participants to focus on single isolated events and evaluate single communication features in human-robot-interaction studies. The two final parts of this section contained references from projects that included context information in social robot scenarios. These references indicate that there may be an opportunity for social robots to use low-detail personality information and adapting their behaviors to it. They also show that it may be beneficial for social robots to use simple physical context information to inform the high-level behavior selection processes.

5 Paper 1 resumé: The Model of Affective Expression Modalities

In the paper titled “A Systematic Comparison of Affective Robot Expression Modalities,” MOAM is introduced. This model is a descriptor and comparison tool for different affective robots. Robots communicate through different interaction modalities, and this paper identifies and focuses on five high-level modalities; morphology, movement and orientation, posture and gestures, audio-based, and anthropomorphic reflection. Figure 5.1 shows the latest revision of the model, and it reflects the amount of emphasis put on specific modalities for each robot. Overall, this works to depict the communicative and affective strengths and weaknesses of an affective robot using a point distribution system to highlight the features of each interaction modality. The model also illustrates how well the robot responds to external stimuli and how aligned its abilities are with the intended work scenario. Each modality is given a rating from 1 to 4 (in the latest revision), and the corresponding section of the model is colored to form a connected diagram. Points are given to each modality according to the following criteria:

0. Point: No interaction modality is present in the robot.
1. Point: The interaction modality aligns with the context.
2. Points: The interaction modality is implemented and aligns with the context, but it only slightly improves the robot’s affective expression ability.
3. Points: The interaction modality is implemented, aligns with the context, and is used to improve the robot’s overall affective expression ability.

4. Points: The interaction modality is implemented, aligns with the context, improves the robot's affective expression ability, and is also used to respond to incoming stimuli.

The limit of 4 points per modality reflects a trade-off between depicting a robot in a high detail level and being able to systematically assess the qualities of affective robots. The downside of using a simplistic model with 4 points is that it may limit the variety in the resulting robot depictions, and the upside is that the model is easy to apply for most robot designers.

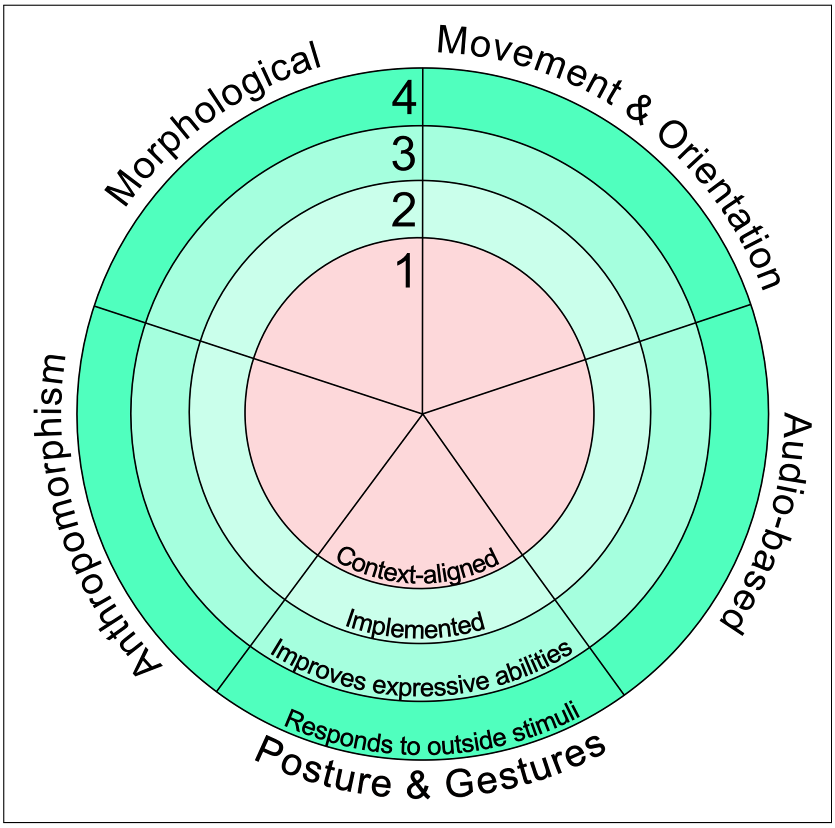

Figure 5.1: The latest revision of the model of affective expression modalities (MOAM) and context alignment. The five slices depict high-level interaction modalities. Each modality can contain one to four possible points that reflects the extend of its implementation.

The maximum points for each interaction modality depicts a robot's ability to react to stimuli in the vicinity using the features from that specific interaction modality. For instance, for the movement and orientation modality, this can be as simple as the robot adjusting its orientation toward a source of audio in its environment. Furthermore, each created MOAM model should be viewed in light of the robot's intended work context and designated task. The inner-circle indicates the robot's alignment to the contextual requirements in any scenario. Each working context introduces specific demands on the interaction modalities and changing the contexts may either decrease or increase the effectiveness of a robot's communication abilities.

5.1 Objectives

The project reported in this paper had the following research objectives:

- To define objective measures for the affective abilities of robots.
- To develop a systematic comparison tool applicable to any robot.
- To use such a tool to gain an overview of the current field of affective robot research.

With the diverse research field in which each project investigates and develops individual communication features for affective robots, this project aimed at defining simple traits for each of these affective robots which facilitated a discussion on and comparison of their affective features. Defining such traits also created an opportunity to gain an overview of the general direction of the research field as a whole to find and highlight previously unused potential research directions for the pursuit of further optimizations.

5.2 Findings

In this project, the model was used to analyze 39 different affective robots. This was done by gathering information from all available research papers regarding each robot and its interaction modalities. With this information, we could distribute points for the modalities in accordance with the MOAM criteria. The results showed that 15% of the robots had points distributed for all interaction modality slices, and 85% had a single modality or more with 0 points. This indicates that most research project robots are designed for tasks that do not require one or more communication outlets. Furthermore, 25.6% of the included robots had a 0 rating for more than three interaction modalities. For the robots that had at least a single modality with 0 points assigned, the average number of modalities was 1.71. This indicates that a large part of these robots was created with a focus on few interaction modalities.

5.3 Contribution

MOAM contributes to the research field of affective robotics by providing a quick overview of the strengths and possible weaknesses of any robot that interacts with humans. This may be useful in the following ways:

- As a checklist for consideration in the design phase of the robot.
- As a comparison tool between already constructed robots.
- As a tool for evaluating opportunities for improvements in an existing robot.

Although the model is simplistic, the simplicity may work in its favor when used in the design phase of social robots. The simple structure that emphasizes five interaction modalities to consider can work as a checklist if the goal is to amplify a robot's affective impact. A more detailed model may provide more direct design directions, but the high detail level often makes it time-consuming or cumbersome to use. A simple model that may work solely as a checklist, has a high probability of being adopted by engineers and being used apart from academic purposes. The same argument can be made about why the tool can be viable: it can provide an overview of a larger range of robots. A more time-consuming analysis of several robots may be infeasible.

The model is relevant as a way to pinpoint opportunities for the further development of the existing robots. In some cases, it can be argued that furthering the development of an existing interaction modality will have little impact on a robot's overall expression ability while adding simple features from another modality can drastically improve it. Such an impact was investigated in our 2019 project as discussed below [46].

The result of creating and using MOAM on various affective research robots indicates that there is an under-researched opportunity in exploring the synergies of using multiple interaction modalities for robots to communicate. This informed the project and led it to include the principle below in the research going forward.

- Multiple interaction modalities should be utilized when designing affective social robots.

This strategy is a departure from the mainstream affective-robot research which tends to focus on a single communication measure to improve human-robot interactions. As previously stated, this can be a viable strategy to ensure that the research experiments obtain clear results from isolated subjects. However, this paper discusses the possibility that it is also limiting the research from gaining insight into the synergies that may arise from combining several modalities.

5.4 Implications

The findings of this paper inspired us to investigate the impact of improving the affective communication abilities of robots with undeveloped modalities. In our 2019 paper, we augmented the audio-based expressions modalities of a robot to strengthen how it was perceived when interacting with humans [46]. In that project, the MOAM model was used on the robot and we identified a key area to improve that would fit the contextual demands of the interaction. The robot consisted of a soft robotic arm and we altered its overall expression by implementing communication features from the audio-based interaction modality. The system used sounds that fit the existing audio originating from the robot. This mitigated the noise occurring from the robot's pneumatic systems. When interacting with the human research project participants, the robot was perceived as being significantly more curious, and happy, and less angry when augmented by artificial audio. The results highlight the possibility of identifying the unused potential of interaction modalities which the MOAM model can aid with.

The model proposed in our 2019 paper has been further developed and used at various stages throughout the current project. The latest version expands the importance of context alignment and can be seen in Figure 5.1. While the inner ring of the initial model depicted whether or not the modality was implemented, the new model places context alignment as the initial entry barrier for each modality. This means that to gain any points in a modality slice, the implemented features of the robot must meet the contextual demands of the robot's work environment. For instance, a small-armed gesturing robot may have the best and most intricate gesturing abilities but if the work context is on top of a large building, nobody will be able to see its small-armed gestures. The work context and intended role of a robot are important as they can limit the number of distributed points for each modality. For instance, for Paro the therapeutic seal, there are no points assigned for movement and orientation. This makes sense as its main therapeutic task is to stay still on the user's lap [61]. This explains why the context is important when creating MOAM models and why different models should be compared considering the roles intended for the robot.

6 Paper 2 resumé: Coordinating robot reactions

In our research paper entitled "On the Causality between Affective Impact and Coordinated Human-Robot Reactions" the focus area was the coordination of events shared between the interacting robots and humans. We investigated the implications of a human and a robot both experiencing and reacting to an event during an interaction. This was investigated with regard to the changes in how the human perceives the robot, and the changes directly related to the timing of actions and reactions initiated by a robot. Knowing exactly when to use specific behaviors and how precisely to respond to human input may amplify how a robot's intentions are perceived. This paper investigated this effect in both cooperative and conflict scenarios.

6.1 Objectives

The research objectives of this paper are shown below.

- To investigate if shared events impact how robots are perceived in an interaction
- To measure how the specific timings of a reaction alter how a robot is perceived
- To determine how to improve a robot's affective impact through delayed responses

6.2 Experimental work details

To attain these aforementioned research objectives, two different test setups were created. The first test isolated and measured the reaction element of affective robot expressions. The second test investigated the effects of delaying the reactions of a robot in a physical conflict scenario with a human participant.

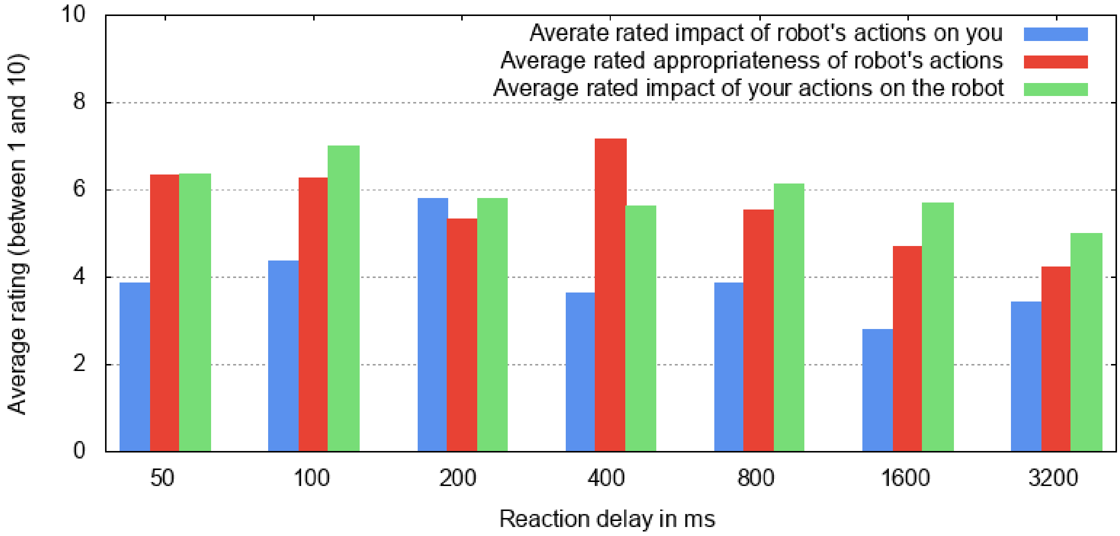

Figure 6.1: The average ratings for each question asked in the experiment grouped by the delay time in milliseconds. The participants were asked about the perceived impact on themselves (blue), the rated appropriateness of the robot's actions (red), and the perceived impact made by the participants on the robot (green).

The initial test was performed with two different groups of human participants experiencing a scenario involving three small robots residing in a small (2x3m) arena. One by one, the participants in the first group were asked to push a red button which triggered a loud explosion sound. In this setup, the participants and the robots experienced the events simultaneously and the robots reacted to the event in coordination with the participants. The participants in the second group were also interacting with the robots one at a time. However, with this group, the explosion sound still happened but the robots did not react to the sound in coordination with the participant. Instead, their reactions were initiated at random intervals.

While the first experiment focused on the effects of coordination versus non-coordination, the second test investigated the exact timing of a robot's reaction. This test focused on the immediate delay in the initiation of a response from a robot after an event. It included a custom-designed robot that could react with greater precision than the robots that were used in the first experiment. The participants interacted with the robot one at a time by hitting it as can be seen in Figure 6.2. By letting the participants hit the robot we were able to establish the exact start time of the event. The robot responded using its eyes and audio. The experiment investigated various response times by delaying the robot's response with each new group of participants varying from 50 ms to 3600 ms.

6.3 Findings

For the initial experiment, we asked the participants to rate the robot's level of arousal. The difference between the perceived arousal level for the two test groups was significant at $p<.05$. When the participants shared a reaction with the robots they perceived the latter as conveying a higher arousal level. This indicates that there is a relation between coordinating a robot's response and how we humans perceive the robot's affective state.

For the second experiment, we asked the participants to rate the perceived impact of the robot's reaction. The average ratings for all the questions per delay time can be seen in Figure 6.1. The results indicated a tendency in the data that the preferred reaction time of the robot was approximately 200 ms. As the human reaction time is approximately 250 ms, this indicates that the participants preferred that the robots reacted with near-human-like response times. The experiments also included questions on how the participants would rate the impact of the response time on the robot (as opposed to the impact on the participants themselves). The response time with the highest rated impact on the robots was slightly faster at approximately 100 ms.

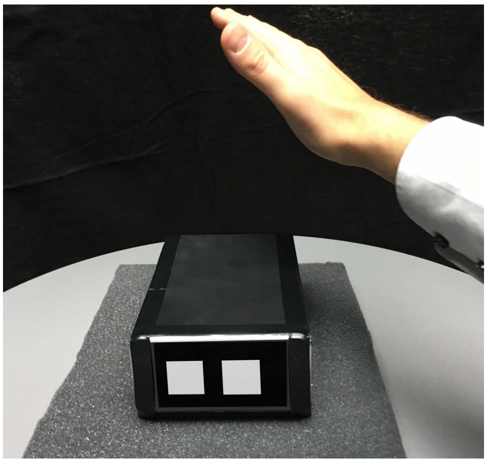

Figure 6.2: The instructional image from the second test which included a physical interaction between the robot and the participating humans.

6.4 Implications

The results obtained from this research project are directly applicable in most robot projects. They show that the events that occur as we humans interact with robots change how we perceive the robots. How we make the robots respond to these events can either increase or decrease the effect of what they are trying to convey. The results show that:

- How a robot is perceived by a human can be influenced by the events they share.
- If the aim is for the robot to make a big impression on the participants, it must react with near-human reaction times.
- If the aim is for the test participants to feel that they made a big impression on the robot, the robot can benefit from reacting with a slightly lesser delay. (˜100 ms faster).

In a condensed format, the findings of this research project helped in forming the suggested design principle shown below.

- Interacting is reacting; the temporal aspects of an interaction can alter the affective impact of robots.

The results of this paper complement previous findings indicating that humans prefer robots that move at human-like speeds in a cooperative setting [168]. Pan et al. 2019 used humanoid robots in such a setting and this project extended these results that they obtained by using a non-humanoid robot and a higher-intensity conflict scenario. It should also be mentioned that the results of this paper are context-dependent and there is no basis for a conclusion that the findings apply to other contexts. For instance, there is nothing that supports a decision to slow down the movement of the robotic arms at factories to human-like speeds. The fact that these results depend on specific contexts both for the task and for the physical work environment should motivate future research to focus on gaining insight into context awareness in human-robot interaction scenarios.

7

Paper 3 resumé: Physical context awareness

In “Adaptable Context-Based Behavior Selection in Autonomous Robots” how to facilitate context-awareness using only simple sensor input was investigated [44]. This research direction was a departure from the main strategy of the previously published papers whose results were highly context dependent. Each interaction modality in MOAM introduced in Paper 1 and the expressions utilized in Paper 2 all had to be optimized for the specifics of the physical environment. For instance, a soft-spoken robot communicates best within close proximity of anyone interacting with it. Robots both in and outside academia often come with behaviors preconfigured for a generic work context in which the robot designers envision the robots to be used. Such robots may have a large variety of behaviors, but they will often not change as the context changes. For instance, the communicative features of a robot will remain the same regardless of the physical circumstances of the robot's task environment such as the noise level, and the distance to the user. If a robot's interaction modalities do not align with the context the robot may be unable to communicate which underscores the importance of gaining (even simplistic) information on the environment. Although there are many ways of gathering information on the physical context, this project focused on measuring and using the simple attainable context information provided by the sensors present in most social robots. We hypothesized that although the gathered context information was simple it would be sufficient to guide the robot toward a general direction with regard to suitable behaviors for its current environment. For instance, a small physical space may inform the robot that large dance moves are not suitable for the current physical context. In this project, we created a system that learns a proper prioritization between predefined discrete behaviors to best align with the different contexts. However, as the number of robots' potential work contexts grows, it may not be feasible for robots to visit and gather details on the behaviors that fit all of them. For that reason, the project also focused on investigating data structures that allow robots to generalize the input and interpolate among the different behavior prioritizations with regard to the measured attributes. This would allow a robot to learn the best fitting behavior for two contexts and estimate a fitting behavior for a context with attributes lying between the two.

7.1 Objective

With this paper, we focused on the research objectives shown below.

- To investigate the feasibility of using low-resolution context information to drive behavior selection in social robots.
- To develop a context representation system that generalizes the gathered data and facilitates predisposed prioritization of behaviors for previously unvisited physical contexts.

7.2 Experimental work details

In this paper, we created and tested a context-aware robot. The robot was not developed to function within a predefined work context but adapted its behaviors as it explored new contexts and interacted with humans. The robot would start by exploring different physical environments. Using a single touch sensor, the robot would attempt to determine the physical room dimensions. This was achieved by driving in a random direction three times and calculating the average time-of-drive number. There are more optimal ways of obtaining the estimated size of the room, but the simplicity of this approach has a purpose: it demands little processing power and requires only a simple contact sensor present in most off-the-shelf robots.

Once the physical attribute(s) of the room was found, the robot would attempt to match the current context with the most similar context in a topography of nodes. Each node would represent a single context. Each time the robot visits a new context the robot would try to cluster nodes that shared similar context measurements. Each node also contained the full set of behaviors available for the robot in prioritized order. The behaviors were similar in that they all consisted of physical movements, facial expressions, gestures and audio. However, the behaviors differed in intensity level. The least intensive behaviors consisted of subtle movements, small gestures, and low-volume audio. The most intensive behaviors, on the other hand, used high-speed movements, large gestures, and high-volume audio.

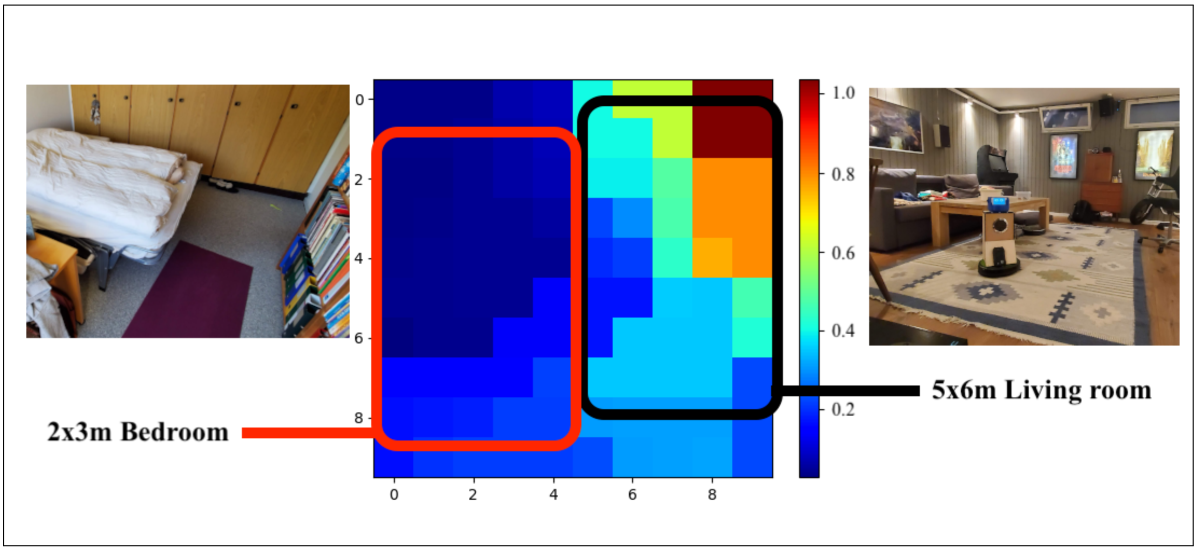

Figure 7.1: The visualized context-map following the initial context exploration. Each square represents a single contexts containing an individual behavior prioritization.

The robot would visit new contexts and interact with humans to learn the best-fitting behavior. With each interaction, it would alter the prioritization for the current context node. It would also update the surrounding nodes (with a lower learning rate). With such an update scheme, the robot tried to establish what worked best for the current context, but also what might be the best fit for similar contexts. The visualized map created after the initial exploration of two different contexts can be seen in Figure 7.1. In our experiments, the robot was tested on six human participants who interacted with it in 72 individual encounters. We tested the robot in two different contexts; in a large physical space and in a small physical space.

7.3 Findings

The initial findings showed that the robot was able to distinguish between the different contexts and that it managed to prioritize its behaviors in accordance with the different context requirements. We hypothesized that high-intensity behaviors would be preferred in the large physical spaces while the lesser-intensity movements would be preferred in the small physical spaces. This was confirmed in the experiments in which the participants, to a significant degree, selected a higher-intensity behavior for the larger space and the opposite for the small space. The robot learned from these interactions and the same result was evident in the behavior prioritization created by the robot during the experiments.

7.4 Implications

The approach to context awareness presented in the paper is a departure from the approaches presented in most affective-robot research papers because it embraces the use of simplistic context knowledge gathered with simple sensors. It may be argued that the results only confirm common intuition about the proximity of robots and the preferred amplitude of their movements and gestures, but the created behavior prioritization is immediately applicable in a large variety of robots. That the map of contexts and the resulting behavior prioritizations were created using only a simple sensor made the approach viable for use in almost every robot project. This created the foundation for the suggested design principle shown below.

- Basic context information may be useful for informing affective behaviors.

From very simple and limited sensor inputs the robot system gained context awareness and used the information to create a simple context-aligned behavior prioritization. In our experiment, the robot behaviors were simplistic and limited to four discrete behaviors. This experimental design was chosen to make the tests more feasible within a limited number of interactions. In real-life scenarios, a robot can prioritize among a much larger set of different behaviors. As the robot's context representation updates the prioritizations not only for the current contexts but also for neighboring similar contexts it allows the robot to visit unexplored physical environments and pre-estimate the best possible behavior to use in that context.

This project also investigated how to mitigate the problems of self-reporting bias, which often occurs when conducting experiments in social robotics using post-experiment questionnaires. This was done by minimizing the time interval between the experiment and the user evaluation. We did this by moving the questions posed to the participants from a post-experiment questionnaire to a central part of the human-robot interaction, in which the robot would ask the questions during the experiment. This approach was used in an attempt to capture the users' immediate sensations and thoughts rather than only obtaining their post-experiment evaluation. We also included non-verbal user measure points to support our findings as a way to circumvent the self-reporting bias.

8 Paper 4 resumé: User group adaptation

The title of the final included paper is “A Minimalistic Approach to the User Group Adaptation of Robot Behaviors using Movement and Speech Analysis” [45]. The paper investigated the feasibility of gathering contextual information from the participants during a minimal set of interactions. The information was gathered for the purpose of adapting a robot’s behavior to the preferences of individual persons and different user groups based on their speech and movement characteristics, as an indicator of an estimated level of extroversion. As a personality trait, extroversion has recently caught the attention of roboticists as it has proven to influence the preference of robot behaviors and how much humans tend to anthropomorphize robots [169, 170]. Extroversion was defined by Carl Jung in the early 1920s but was also presented in Eysenck 1965 as one of the main attributes in the “Personality Inventory” [146, 155, 156, 171]. In this project, we defined the measurable attributes in a human-robot interactions that could distinguish them across several contexts, and potentially indicate a general extraversion level of humans. Furthermore, we used this information for informing high-level behavior prioritization. It was important that the data could be gathered within the interaction and that the required interaction was only a short one.

The defined attributes were based on the speech and movement characteristics of the participants as these relate to how energetic, talkative, and outgoing they were perceived. These perceptions could indicate their level of extroversion. As discussed in the paper, there has been progress in using both audio and movement patterns to define user characteristics. This project focused on obtaining similar results but with a simpler sensor setup and within a short interaction duration. The project also investigated how these measured attributes are impacted when the physical contexts of the interaction are changed, and determined the viability of distinguishing individual users across different physical locations based on the selected attributes.

8.1 Objective

With this project we focused on the research objectives stated below.

- To determine the feasibility of using simple readily available data to distinguish individual users.
- To determine the feasibility of generalizing these attributes to form user groups and of adapting robot behaviors to them.
- To determine the impact of interacting across several contexts on the measurable attributes.

8.2 Experimental work details

The experimental setup included a custom-developed control system in a robot equipped with audio and camera sensors that enabled it to gather the necessary data from its interaction with humans. The data were gathered as a part of the experiments performed in our 2021 project [44]. To test the user distinction abilities of our system it was necessary to include multiple interactions with the same participants. All in all, the robot interacted with six human participants in 36 interactions across two different physical contexts. The robot interacted with the human participants by asking them questions and by reacting to their answers with a scripted response. As the aim of the project was to determine the feasibility of using a minimal dataset and a simplistic sensor setup to distinguish individual users, the main data were gathered within the first two questions in the interaction. The environment of the interaction, how the participants moved during the interaction, and the speech patterns in the participants' answers was the data that were used in the further analysis. In our experiment, the robot also asked the participants to select one of two behaviors and proceeded to show them two behaviors similar to the experiment in our 2021 project [44]. The selected preference was noted and these data were used to investigate the relation between the preference and user groups based on the measured personality attributes. The test participants interacted with the robot one after another with each interaction lasting about 5 minutes. With each physical context, the participants interacted with the robot three times.

8.3 Findings

The findings of this project revealed that the robot could successfully distinguish individuals in a single context using the measured attributes. This was evident in the data as the measured data points per individual were more closely clustered together than the overall standard deviation. The results for distinguishing the participants across the two physical contexts can be seen in Figure 8.1. As the graph shows, the measurements of the participants' personal traits are closer to each other in each physical context than in a different context. This means that the robot will be able to distinguish the users individually in each separate context but will not be able to recognize users across multiple contexts.

The foregoing can be explained by the fact that the selected attributes are insufficient to give proper insight into the different user characteristics. The minimalistic setup could also have limited the variety in the data for each participant as two questions could not have left much time for the participants to stand out. The failure to distinguish the participants across different contexts can also be attributed to the fact that people may talk and move slightly differently in each context. The differences in the physical dimensions of each test environment might have enticed the participants to move more or less as the space around them increased or decreased.

The data regarding behavior preference were analyzed in terms of the correlation between the users' behavior preference and the possible similarity of the user's speech and movement attributes. To adapt the behavior of the robot to that of the user groups on the basis of these attributes, there should be a similarity between the behaviors chosen by the participants and their individual characteristics. However, beyond a few coincidental similarities, the data gathered in the experiments were insufficient to establish a significant correlation between the two.

8.4 Implications

The paper presents a method of gathering contextual information from the participants within the first moments of an interactions. The sensors that were used to gather the information are present in most social robots which makes the results immediately applicable in many robot projects. That the robot was able to distinguish individual users only within isolated contexts, however, indicates that there is room for improvements. The approach may be strengthened by combining the speech and movement attributes with other systems that inform the robot about the current physical context. For instance, if the robot can distinguish the users in each context, gaining information on each current context can aid the robot in distinguishing the users across several contexts.

It may be argued that defining user groups based on personality information gathered within two sentences of an interaction is infeasible or maybe too ambitious. The Eysenck personality index uses more than 50 questions to estimate the same. However, using a limited number of questions makes the system more practical and usable in a real-life work scenario. The findings of this project highlight the strengths of using multiple sources of easily attainable information to provide knowledge on the current context, and they informed the creation of the suggested design principle shown below.

- Combining multiple sources of low-complexity context information can increase the contextual insight.

Many previous research projects focused on various aspects of a user's personality, but the approaches that were used in such projects often require an elaborate sensor setup or the use of pre-experiment questionnaires to gather information on personality factors, such as the Five-Factor Model or the Personality Inventory. This is not feasible for any robot in a real-life scenario that has to function without the information gathered through a questionnaire administered before every interaction. By tolerating the impreciseness and uncertainty of the result, robot designers can use this approach to increase robots' context awareness. They can either choose to limit a robots' decisions on the basis on this information or enable the robot to combine such information with other easily attainable information to gain a richer contextual understanding of the situation.

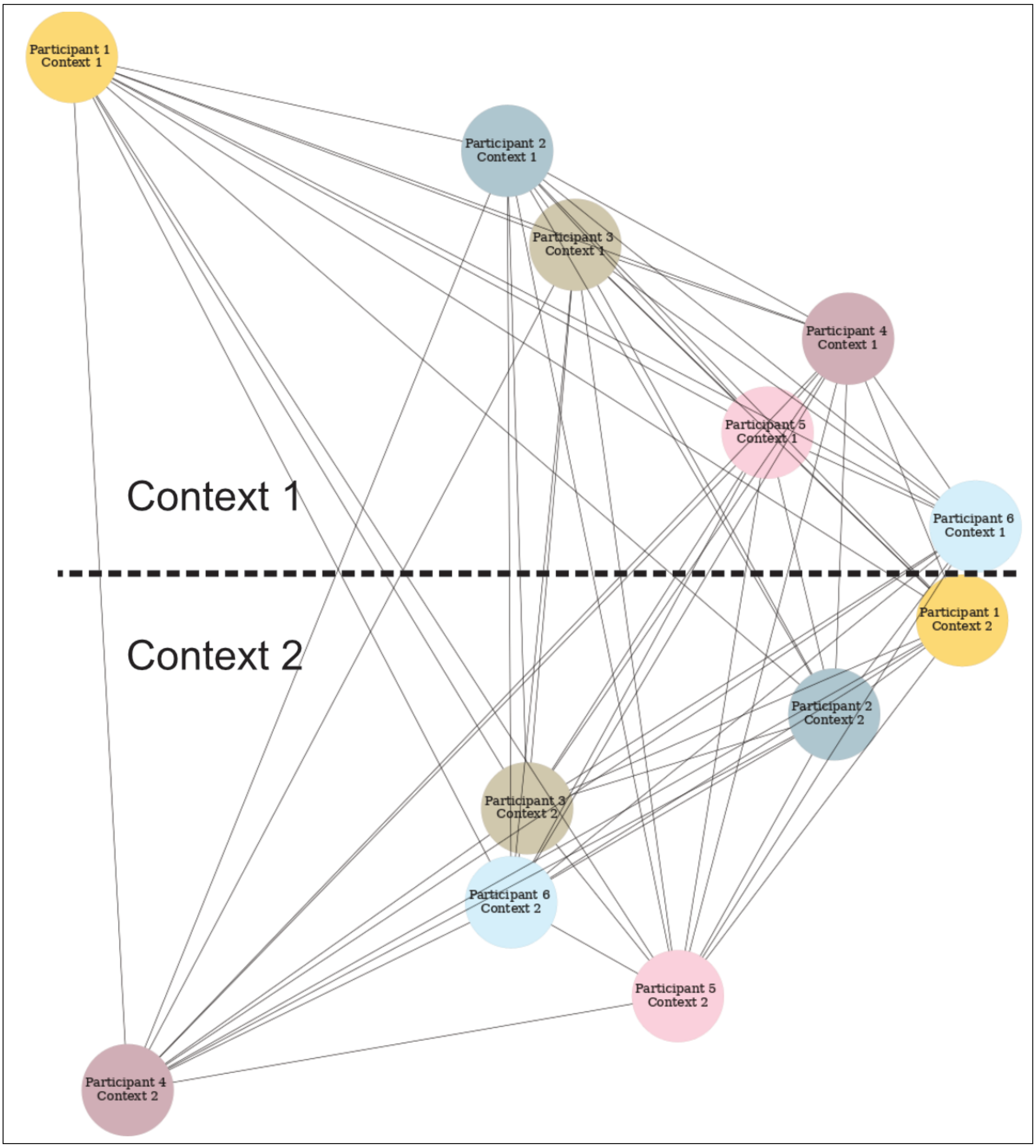

Figure 8.1: Each color represents a participant, and the two circles for each participant represents their measurements in the two different context. The edge between circles represent the Euclidean distance between the vectors with their measurements. The measurements of personal traits for each participant are within closer distance to each other in each physical context than to themselves in the other context. This means that the robot would be able to distinguish the users in each separate context but not across multiple contexts.

9 Discussion

This dissertation introduces seven design principles that can help robot developers build robots with a stronger affective impact, and with the ability to communicate using all available means of interaction. The design principles are not strict rules but guidelines that may be used to inform the designs of social robots. They are also meant to serve as a tool for consideration when conducting experiments involving human-robot interactions.

The principles are as follows:

- Anthropomorphism may be emphasized to increase the affective impact for robots.
- Perceived complex emotions can emerge from the combination of simplistic behaviors.
- When and how humans perceive robots and evaluate robot behaviors are factors that are difficult to control.
- Multiple interaction modalities should be utilized when designing affective social robots.
- Interacting is reacting; the temporal aspects of an interaction can alter the affective impact of robots.
- Basic context information may be useful for informing affective behaviors.
- Combining multiple sources of low-complexity context information can increase the contextual insight.

It can be argued that encompassing all the above design principles in each robot project may raise the overall complexity of the project. It can also be argued that it may be difficult to transform the principles into practical design directions in a robot project. This section will discuss each design principle considering the following points:

1. How we attempted to incorporate the design principle.

2. How the principle may be applied to different scenarios.

The principles will be discussed in the following sections, with a focus on the engineering aspects of using each principle and through the hypothetical evaluation of the applicability of each principle.

9.1 Using anthropomorphism

- Anthropomorphism may be emphasized to increase the affective impact for robots.

As discussed in Fussel et al. 2008 and as shown by the 1944 Heider and Simmel experiment, it is difficult to avoid anthropomorphic interpretations of a robot's behaviors [8, 29]. We found that no matter how simple the robot presented in an experiment is, the users will project different human emotions and intelligent intentions onto the robot. With anthropomorphism being an integral part of the interaction it is evident that robot designers have to be aware of how it may influence the perceived intentions conveyed by any robot. It thus becomes vital in affective-robot research to investigate what stimuli either add to or subtract from a potential anthropomorphic interpretation. In our project, we investigated how to strengthen anthropomorphic interpretations of different robot behaviors through the synergy between the behaviors and engineering aspects of affective robots. Utilizing anthropomorphism does not necessarily mean replicating human features. Although some features of the robots we developed were biologically inspired and human like, we generally avoided using humanoid robots. Instead, we focused on the behaviors and reactions of non-humanoid robots. We also emphasized anthropomorphism by using the contexts and narratives of the robot's actions in the experiments. This was evident in our 2020 project, in which a high-intensity conflict situation intensified the participants' feeling of empathy toward the robot [43].

It is difficult to directly use anthropomorphism. The interpretation of a robot's behaviors cannot be forced, is never universal, and differs from person to person. For instance, during our experiments, some participants would gladly hit our robot, stating that it was merely a box, while other participants did not want to hit it or just felt very sorry for the robot while hitting it. One possible way of emphasizing the effects of anthropomorphic interpretations of a robot's actions is to flip it around and to try to avoid any feature that would subtract from it. We did this in our experiments by aiming to entice anthropomorphic interpretations with the way we presented the experiments and robots to the test participants. We refrained from referring to the robot as "it" and from explaining the inner workings of a robot during the experiments as so not to prime the participants into thinking of the robot as an object.

- Perceived complex emotions can emerge from the combination of simplistic behaviors.

Anthropomorphic interpretations of robots' actions can be useful with their ability to attract a finite interpretation even from basic cues in an interaction. Humans often want to project intelligent intentions onto a robot even when they see only a moving box or flickering lights. This may be used to reflect a more intelligent robot behavior through simple communication means. In other words, we do not always have to be specific when we design expressive behaviors as people will project an interpretation onto even simple robot behaviors. When the robots got stuck on the carpet of the arena in our 2020 project, the participants quickly concluded that the robots were just resting instead of perceiving it as an actual error [43].

Applied to robot projects in general, this means that we do not necessarily have to copy how humans portray complex emotions. The mix of context and simple robot behaviors will often result in a complex interpretation. Similar to how humans interpret each other's actions in light of the current situation, the context influences how the robot is perceived. In our 2020 project, we combined simplistic behaviors and utilized contextual events to make our robot seem remorseful and apologetic [36]. All the behaviors that were used to convey such specific complex emotions were simplistic and consisted of basic audio, simple movements, and gestures. The actions of the participants and how they interacted with the robot in the experiment augmented how these simple actions were perceived. Furthermore, if robot behaviors are initiated immediately after an outside event occurs, the robot will often be perceived as having an opinion of the event. This happens often even though this opinion is actually formed by and exists only in the minds of the observers.

- When and how humans perceive robots and evaluate robot behaviors are factors that are difficult to control.

Some of the references in this dissertation indicated that we humans perceive and evaluate our current context indefinitely and subconsciously. When we in academia design robotics experiments including human participants, we sometimes rely on a definite start and finite end to the experiment. This is often followed by questions posed to the participants. We also often rely on the scientific conventions of isolation of the subject under investigation and isolation of the test parameters. Such design of the experimental setup conflicts with an evaluation process that begins before and continues after the experiment, a process that also incorporates contextual inputs from outside the planned experiment parameters. In our project, we acknowledged that the evaluation happens outside the experiment and we countered the effect of this by making sure that our robots were complete in our experiments, in the sense that no loose wires, cogs, or inner machinery were visible. We covered any unwanted features that could influence the interaction and how the participants would interpret our robots. This included paying attention to our robot's noises. We included audio to cover the noise from any moving parts for various actions of our robots. This meant developing appropriate sounds for gesturing or for moving from point A to point B. We also created a consistent test environment, in which the participants and the robots interacted with each other, to exclude any unwanted outside input. For that purpose, we built separate physical rooms inside the laboratory so that the experiments could take place with one participant at a time. In our last two projects, we also moved the time that we asked the participants the questions to within the experiments. This was done to avoid any self-reporting bias. We also asked the participants questions that could highlight something else unforeseen outside the isolated object of the investigation.

We tried to be aware of the interdependencies between the different aspects of an interaction. Figure 9.1 shows the circular dependencies of the investigated subjects. The expression capabilities of the robot are dependent on being attuned to the context and the context can be altered by the robot's actions. The expression abilities of the robot are dependent on how it is perceived by the users, and the users are influenced by the context in which they may either like or dislike the experience. The users may alter the context while the expression abilities of the robot can alter how the users feel. The circular dependencies of these entities also illustrate that it may be challenging to design a robot with a focus solely on one of these aspects as the other two are intrinsic parts of the interaction. Although these focus areas are often investigated

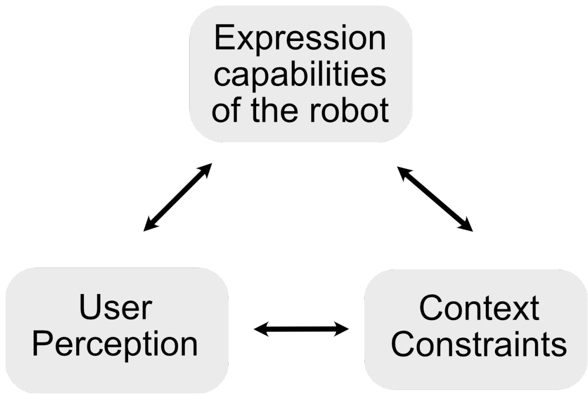

Figure 9.1: The interdependencies between the areas of interest influensing the interaction. The arrows represent overlap between the different areas.

subsequently, in reality, they may influence each other. This may be considered in any discussions on isolated project findings as the interdependencies of these three entities may influence the findings. This does not mean that any finding is incorrect but that any discussion on a finding's application may benefit from considering all three aspects of the interaction.

It may be argued that this design principle relates mostly to how robots are used in laboratory experiments. However, the concept may also benefit robots in other areas outside academia.

Flipping around the examples given in the introduction and focusing on how the humans in the interaction experienced the events can result in the scenario below.

> *The guests are ready to order and the waiter robot approaches the table. The robot is unstable and shakes a bit as it approaches the table. The robot has a small screen that shows its eyes, and there is an area with dead pixels on the right side of it.*

The technical problems of the robot are minor, but the guests may be hesitant to trust it to complete waiting on them. Although the robot's problems may have nothing to do with the robot's functionality, it may still influence the guest's trust in the robot and willingness to interact with it. Outside the technical problems, everything in the context can influence how the robot is perceived. For instance, a crack in the outer shell of a vaccine robot can make people scared of getting vaccinated by the robot. It is beneficial for scientists to acknowledge that the experiment may actually begin prior to it or may continue beyond it. As such, it may be relevant to consider all factors that may influence an interaction. Even though some factors outside the experimental setup may seem technically irrelevant for the robot project, such factors may end up providing the needed information to more fully comprehend the human-robot interaction.

9.2 Utilizing interaction modalities

- Multiple interaction modalities should be utilized when designing affective social robots.

The introduced MOAM contains five different high-level interaction modalities. It may be argued that the low granularity of the model may result in an unspecific description of the intended role of the robots. However, the low granularity also provides some freedom in determining whether and how to implement each modality. The robots that we developed for the experiments in the included projects were designed to create simple implementations of each modality so that the full range of modalities could be covered. This also included implementing different reactions using each modality. This was done to make the robots capable of forming a fitting answer to most of the outside inputs, to make the robots seem more believable, and to strengthen how well they invoked an anthropomorphic interpretation of their actions. Our project also focused on the immediate high-impact improvements of the robot. Rather than implementing advanced or complex solutions for a single expression modality, we implemented simplistic measures to highlight the unused potential of any unimplemented modality.

It may not always be possible or necessary to implement all modalities. The extent of the implementation largely depends on the work context and the intended role of the robot. MOAM can work to highlight the areas that will provide the biggest affective return of investment when altering a robot. Many research projects aim to improve a robot's existing features to improve its affective communicative capabilities. What MOAM is trying to convey is that there is a simple alternative to improving a robot even further. By adding simple implementations of previously unused modalities, it is possible to radically change how a robot is perceived. In our 2019 project, we found that adding simple audio capabilities to a soft robot made the project participants who interacted with it perceive it as being happier, sadder, or more playful. However, the participants perceived the robot, their impressions of it were stronger [46]. This shows that a diverse set of simple implementations of different interaction modalities offers a strong overall affective impact and that improving the missing areas in the model may be a good investment.

Anthropomorphic interpretations are a vital part of the interactions between humans and robots and may be impossible to avoid. Strengthening how much humans project intent and motives onto robots can make it easier for robots to communicate and convey affective information. However, that is difficult to control as roboticists can only control the expression of a robot and not its interpretation. MOAM emphasizes robot's reactions to outside inputs using different modalities to form the reactions, and this may strengthen how people perceive the robots. For instance, a robot that does not move even when one is physically interacting with it may not be perceived as a living entity while a robot that shows simple reactions to various input types may be perceived as real and as an entity that comprehends the situation.

Once again turning around the example from the restaurant and looking at it from the perspective of the humans in the interaction may result in the scenario below.

> *A restaurant robot is waiting at a table and the guests are going through their orders one by one. The guests find the robot fascinating and trust it in the given situation. One of the guests, however, spills a glass of water. The robot continues to ask about the order. The guests are trying to stop the robot by shouting "Stop!," doing wild gestures, and tapping the robot on the head, but to no avail. The robot understands only specific spoken commands and continues unfazed by the events.*

It makes sense to enable robots to understand and respond to multiple kinds of input. The lack of response by the robot in the above scenario may result in the guests' lack of trust in it. This is also important for how the robot communicates and is similar to humans. We all communicate slightly differently, some people rely heavily on gesturing while others express themselves through variations in the amplitude of their voice. Any future breed of robots should be able to both use and comprehend such variations.

9.3 Applying event coordination

- Interacting is reacting; the temporal aspects of an interaction can alter the affective impact of robots.

Both the included references and the experiments we conducted indicated that a robot's reactions to outside events influence how the robot is perceived. In our project, we investigated the coordination and timing of reactions for robots and found that these had an effect on their affective impact. This required that the robots we used could react fast and control their reaction timing within milliseconds. However, not all of the robots we developed could react with a high level of precision. For some of the robots, their processing time of the measured input could make them appear a bit slow. We countered this effect from a software design perspective by doing away with the input processing that was not important to the interaction.

We aimed at making robots that could react using different kinds of physical and non-physical communication methods. We embarked on this in our 2020 project, to ensure that the robots' reactions would be well conveyed and to investigate how the hierarchy of modalities can influence how robots are perceived [36]. We found indications that what mattered most was not the specific type of expression modality used in the robot's reaction but that the robot reacted through a broad variety of interaction modalities. This shows the importance of making robots understand and react to their surroundings.

Although we used a robot that could react with high precision to investigate these aforementioned concepts, the same technical setup is not needed to apply our findings in other robots. We found that experiencing the same events as humans do and reacting to such events when humans react to these strengthen the affective impact of robots. This may indicate that there may be an unused potential in enabling robots to understand different kinds of stimuli and react to them. This can make them seem more believable and can strengthen their bond with any human they would interact with.

Let us once again look at a restaurant example from the guests' point of view.

> *As the robot is taking an order from a guest another robot in the restaurant drops a plate. The robot jitters, looks around, and comments on it to lighten the mood and ease the tension.*

Interaction is reacting, meaning the immediate actions a robot carries out after an event occurs are important and can set the tone for how the robot is perceived throughout the rest of the conversation. This also applies to turn-taking in human-robot conversations in which embodied and voiced reactions are used both to communicate engagement and to convey a wish to initiate a sentence. In our projects, we enabled a robot to react while the humans were speaking to convey that the robot was listening and was still engaged in the conversation. Timing such reactions can also strengthen this effect. Reacting too slowly may seem inappropriate, and reacting too fast may cut off the speaker while still talking. Robots are often able to react faster than humans but this may not always be the best option. It may even often be beneficial to slow down the working speeds of social robots. This is especially vital for cooperative tasks in which the humans and robots have to work together. Furthermore, slowing down the robot's response time can allow more time for the robot to process the input and select the optimal response for the given situation.

The results of our experiments indicated that the perceived target of an emotional impact in an interaction may differ with different reaction speeds. However, there may be mechanical limitations to the extent to which a robot can guarantee a fast reaction. In our project, we had to rely on audio and video as communication methods rather than on gesturing or movements to facilitate subhuman reaction speeds. This was because our hardware was too slow to provide a physical response. There may be an opportunity to investigate the synergy among the different interaction modalities in reactions when one of them is applied with faster timing than the others. For instance, if a robot cannot react precisely enough using gestures, it can perhaps benefit from reacting swiftly using audio and then adding a physical reaction more slowly with gestures or movement. This may also include an investigation of the timing between the two as delaying the second physical gesture too much can make the robot seem odd. Also, with better hardware, there is an under-researched opportunity to further investigate reaction delays to control the perceived target of the emotional impact. This may be used in cases where the goal is to invoke empathy for the robots.

9.4 Attaining low-detailed context awareness

- Basic context information may be useful for informing affective behaviors.

The physical context can influence how the actions of robots are perceived. An action that seems fitting in one context may not work in another context or may mean something other than what is intended. With the initial experiments and robots developed in this project, the physical context and work scenario of the robots were highly influential on the shape and size of most of the robots' communicative features. This meant that the robots were constructed for the specific targeted physical work environments and that their interaction modalities were aligned with such specific contexts. For instance, we enabled a robot to move on a level surface and attuned the volume to the specific noise levels in the laboratory. We also adjusted the targeted distance between the humans and the robots in the interaction. Adhering to specific physical requirements meant that our robot would mainly work under the conditions of our laboratory or in another context with similar physical attributes.

In our project, we created experiments that directly focused on the physical contexts and their impact on behavior prioritization for social robots, but the notion of contextual awareness also influenced other projects. We attempted to isolate the tests as much as possible by constructing a "neutral" test room for each of the experiments. It may be argued that a neutral context does not exist, but we aimed to exclude as many outside inputs as possible that could pull the focus away from the experiment and to provide consistent and similar settings for each experiment. We also used adaptive behaviors that allowed the robots to react to inputs with the same intensity as the received inputs. For instance, we adjusted the volume level of the robot so it would match the measured speaking intensity of the humans in the interaction.

The physical conditions allow robots to successfully communicate. Adapting to the contexts can make robots function in multiple different environments. Successful current consumer robots such as Pepper come pre-equipped with morphology, communication features, and locomotion systems designed for specific working contexts [54]. These features work great in the targeted contexts but are difficult to employ in other physical domains. Such predefined robots are also rarely designed to switch between contexts on the fly and adapt as the contextual demands change. We argue that even simple contextual data can be used to make the robot seem believable.

Let us look at another restaurant example as experienced by the guests.

> *The robot is informing the guests about the available entrées. It is a small intimate restaurant and the robot has set the volume level of its voice to match the atmosphere. Another family is seated at the adjacent table. They start arguing loudly. The robot continues to serve the family but raises the volume of its voice to ensure it can be heard.*

The data needed to inform the robot about the physical room dimensions can be gathered using a simple contact sensor, as in our 2021 project, and a noisy environment can be detected using a microphone [44]. It can be argued that the robot will need to verify that the noise did not come from the guests themselves it is waiting on, but the example still shows that even simple, easily attainable information can be used in some capacity to inform the robot's behaviors.

9.5 Combining contextual data sources

- Combining multiple sources of low-complexity context information can increase the contextual insight.

Although the use of simple available data to form an immediate reaction during an interaction may have an advantage for robots, there may be scenarios in which a single source of information does not provide sufficient contextual information on its own. In our project, we had a robot that attempted to use the speech and movement characteristics of humans to distinguish them from each other [45]. The system worked in isolated environments but failed across multiple physical contexts. The project thus suggested that measuring distinctive attributes in each physical context could enable a robot to work across several contexts. Combining simple sources of information in each context can provide a robot with a low-granularity overview of the current situation, one that is both usable and attainable.

With our robot in our 2020 project, we tried to combine many different data sources to gain contextual information on the intentions of each person currently communicating with the robot [36]. In the experiment, we informed our robot of the intensity level of the current mood of a human scolding the robot. Our initial implementation of a system that reacted to loud voices was not sufficient as it also reacted to loud happy sounds. The combination of several data sources, such as the measured voice amplitude, sentiment analysis, and facial expression classification, provided enough contextual information for the robot to determine that the person was scolding it rather than making fun of it (which the participants also tried to do in an attempt to trick the robot into reacting to such). Although these individual sets of information require complex processing, the sensor requirements for gathering them, such as a microphone and a camera, are readily available in most off-the-shelf robots. The individual information sources did not provide enough information, but the combination of several input sources completed the contextual image.

Let us look a final time at a restaurant example from the point of view of the guests.

> *The guests are celebrating a birthday party at the restaurant and the robot waiter approaches them. The robot hears multiple loud noises and sees*

one energetic person moving a lot. The robot decides that several energetic people with loud voices may indicate something about the social context. (e.g., party, birthday). It alters its behavior (and tolerance level) before approaching the table. The guests experience a robot that waits on them attuned to their festive behavior, a robot perceived as being both energetic and like a natural part of the festive atmosphere.

The intelligent behavior in the above scenario emanates from two sources of information: The first piece informs the robot that the voices are loud and that the persons may be in a high-intensity mood. The second piece of information informs the robot that there are multiple people in the vicinity. These contextual cues can be gathered with simple means but may have a great effect on how well the robot comprehends the current social context. Also, as with Russels and Mehrebians PAD space, detecting intensity in combination with the number of people in the vicinity may be a great approach to classifying various social contexts [99]. The key challenge with using the information gained in this approach is deciding on a fitting abstraction level that matches the processed contextual information. For instance, the physical room size may be used to determine the speed of a robot's expressive movements but could be ill-suited for informing a robot on how to invoke empathy in a conversation. Many robot research projects disregard context awareness, arguing that there are too many variables to consider and that the gathered data never provide complete or sufficient context awareness. With this project, we argue that instead of trying to extract a complete overview of the contextual situation for a robot, we should accept that the data will always be limited or imprecise and thus adjust the scope of our application of the information in the current interaction.

10 Conclusion

The main research objective of this project was to determine how we could develop more believable affective robots, and how we could increase such robots' context awareness ability. To attain this objective, we focused on the following sub-objectives: to determine the impact of using different expression abilities of affective robots and of coordinating reactions in an interaction, to determine the feasibility of using immediate and simple physical context information to drive behavior selection, and to determine the viability of using a minimalistic interaction to gain contextual information on the humans interacting with the robot. The investigations were performed through the development of several prototype robots and through experiments in human-robot interactions.

The main conclusion that can be drawn from our overall findings is that to build more believable affective robots, we should start focusing on using a wider spectrum of expression and input-sensing abilities when we design them. The immediate gains from developing a pre-existing single affective feature to further increase the affective expression abilities of robots may evidently follow the law of diminishing returns. There is still plenty of room for improvements, but to create more believable agents, it may be a better investment to focus on utilizing a broad line-up of varied additional affective abilities than to focus on improving individual features.

There is a definite heritage from Rolf Pfeifer's complete agents (which are built with attention to all aspects of autonomous communication) in our departure from mainstream robot development and our findings indicate that we are at the same type of scientific crossroad. There can be a paradigm shift in how robot designers build and test robotics if we disregard some of the conventions that are holding us back. For instance, there is a long tradition in academia of ensuring simplicity in the experiments and decreasing the number of variables that may influence the results. In contrast, we found that the test participants may be influenced by input in each human-robot interaction experiment that is outside the controllable parameters of any interaction. The research and experiments we performed revolved around creating a strong communicative impact by altering how we perceive robots. That is a complex effect to measure through the conventional quantitative scientific methods. We conclude that it may be infeasible to completely isolate the effects of single affective robot communication features. This may call for a more holistically inspired approach to how we conduct experiments in human-robot interactions. The experimental setup may still isolate the subjects, but the approach to capturing the details of any interaction may benefit from including enough information on the events to gain a full perspective of the experiment. This will also allow effective robot projects to include a broader range of affective features in the robots that are used and to allow researchers to build more complete (MOAM) robots.

A shift in perspective is needed to start constructing robots with a greater potential for being used outside academia. As stated by Pfeifer 1996, if we keep building apple-picking robots in a laboratory they will never be able to actually pick apples outside laboratory conditions [17]. If we want robots that can handle social situations outside of human-robot interaction experiments, we need to begin including a fuller skillset of the features that will enable them to function autonomously in a social capacity. The current robots generally lack the skills to be aware of what is going on around them. From our findings, we conclude that robots may gain a stronger affective impact if they react to outside inputs. They also gain a stronger affective impact as they share reactions with their users, and adapt to each user in the same manner as humans do in every conversation they partake in.

It will be difficult for future robots to handle real-life socially complex situations if we do not start giving them a context awareness ability. Context awareness is the natural evolvement of robots that can express complex affective information. When we already have a robot that communicates using multiple modalities and reacts to various kinds of input from each interaction modality, the next step is to stop sending robots into each interaction unaware of what is happening around them. A comprehension of the current context on a human level is extremely difficult to achieve, but we found that context awareness could be made simple and usable as long as the robots use it only to make simple decisions at a fitting abstraction level. We conclude that it is feasible to start using simple, easily attainable sets of context information to improve the context awareness of the robots in most of the current robot projects. The solutions may be as simple as lowering a robot's voice as it gets closer to a user or allowing a robot to increase its travel speed when there are no humans in its trajectory. We acknowledge that human-level complete context awareness is still the next frontier but we should already start using what we can now.

When robots gain awareness of the events occurring in their immediate vicinity, this allows them to be reactive to a wider range of contextual inputs. This can influence how we perceive them in an interaction as the shared coordinated reactions between humans and robots may spark an immediate bond between them. Using simple context attributes to achieve this may facilitate this effect. We envision the next breed of rescue robots as capable of reacting to and acknowledging all the sounds that may scare a frightened human trapped in a tight space and as thereby capable of gaining humans' trust in them.

The simple, easily attainable information is available not only in the physical context but also in the characteristics of the humans with whom robots interact. This information is readable in every interaction and humans use it all the time to adapt their behaviors to the people they interact with. If we encounter someone extraordinarily happy we try to match his or her energy level. If someone seems tired or sad as he or she approaches us, we rarely dance around but try to match his or her behavior to show our support. Such information does not require extensive personal background knowledge on the humans we interact with but may be gathered from the immediate cues in the context of the interaction. Reacting and adapting to such information can make a robot more believable and can increase its viability to function as a companion in a stressful situation or to act in a therapeutic capacity.

We have stated that the concept of anthropomorphic interpretations of robotic behaviors can help robots portray complex emotional scenarios. We found that it is difficult to completely avoid anthropomorphic interpretations of a robot's actions. As a result, we conclude that the best approach to handling the influence of such on human-robot interactions is using anthropomorphism to our advantage. This entails maximizing the robot's affective impact with behaviors that support an anthropomorphic interpretation.

Finally, we conclude that it is the synergy in the combination of all the individual robot abilities that creates a truly convincing social robot that is tolerable to interact with and has the capacity to adapt to each user in a meaningful way. We suggest that encompassing all these responsibilities in one large system may be infeasible, but creating multiple parallel single-responsibility systems with each robot responsible for processing a single piece of context information can make the system feasible and can provide sufficient contextual information.

The main keyword we have worked toward during this project is applicability. We aimed at creating easily replicable robots using off-the-shelf parts with simple controllers, and it is vital that our results are directly usable in most projects. By developing simple affectively impactful robots that understand their immediate context and react to it, we have taken a step toward creating a new breed of social robots that may work autonomously outside academia, and the way to realize such robots is through a paradigm shift toward a future with robots we can all benefit from.

"Go forth, affective robots! Learn about the context and react to it!"

11

Paper 1: A Systematic Comparison of Affective Robot Expression Modalities

A systematic comparison of affective robot expression modalities

Morten Roed Frederiksen[1] and Kasper Stoy[2]

Abstract— **This paper provides a survey of the different means of expression employed by robots, to convey affective state to human recipients. The paper introduces a model of affective means (MOAM) to effectively describe and compare the emphasis on specific means and applies it to the surveyed robots. The model entails viewing the effect of applied means in light of how well the robot responds to external stimuli and with attention to how aligned the robot's means of affective expressions are with the intended working scenario. The model-based survey shows that a majority (85%) of the surveyed robots contain a category with room for additional affective means, and a quarter (25.6%) of the robots use a single or two affective means of expression to convey affective states. The result of the survey indicates there is an under-researched opportunity in exploring synergies between means of affective expression to amplify the overall affective impact of a robot.**

I. Introduction

To improve the way robots interact with humans, the intentions of the robots need to be easy to interpret. This means that the information they convey about their current status and intentions is easily readable and warrants no further need for formal explanations [1]. One way to reach such communicative skills is by enhancing the interaction using affective means of expression. These means could be comprised of the robots appearances, the way they move, how they gesture and pose themselves, how they sound and whether we are familiar with what they portray [2], and lastly how they respond to incoming communication [3], [4].

In contrast to robots, humans use subtle cues such as body language, tone of voice, gestures, and movement in a constant negotiation of affective status through each encounter with each other [5]. Even before the interaction is initiated our posture and general appearance sparks an initial presumption of our current mood and intentions towards the interaction [6], [7]. These affective measures emphasize the messages we want to convey and influence how well they are received.

In addition to being able to express affective status some degree of emotional understanding is also demanded from the robots to improve the interactions with them. Since affective computing and emotional intelligent systems were reintroduced by Picard in 1997, a significant amount of research on the topic has centered on how to measure human affective status [8], [9]. This has yielded successful results using facial recognition [10], electromyography [11], gesture recognition [12], voice patterns [13], and touch measurements [3]. Although the emphasis on how to measure affective changes is relevant for realistic social interactions between humans and robots, this study aims to give an overview of the different means for robots to express affective states and to provide a model for describing and comparing affective systems. Therefore the focus will lie solely on the technological capabilities of the robots to convey emotions.

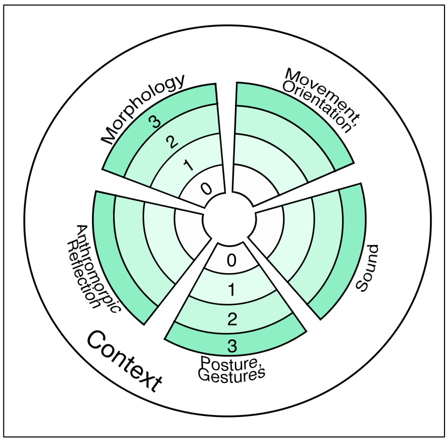

Fig. 1. The model of affective means (MOAM) and context alignment with with 0 to 3 points distributed in each of the included five categories of means. (See the Expression modalities section for further explanation of the systematic point distribution)

A majority of the robots included in this survey were created to test and improve single means of affective expression in isolation from other means. However, there are indications that means can influence each other and distort or increase the emotional impact of any emotional expression on human recipients. Eg. adding music to a scenario might alter a negative mood towards the positive [14]. By exploiting synergies between means, we can emphasize the intention of the robot beyond what is possible with one mean alone.

II. MOAM - Model Of Affective Means

This survey proposes a simple model to facilitate a systematic comparison of affective robots. The aim is to illustrate how much emphasis is placed on specific aspects of the affective means and to work as a tool for robot engineers to employ in the design phases of robot construction. The model divides the available affective means of robots into five high-level categories. The chosen abstraction level is the result of a trade-off between the ability to depict robots in greater detail and the ability to easily compare different robots. The current model favors the last of the two. The

*This work was not supported by any organization

[1]Morten Roed Frederiksen mrof@itu.dk and [2]Kasper Stoy ksty@itu.dk are with the REAL lab at the Computer science department of The IT-University of Copenhagen, Rued Langgaards vej 7, 2300 Copenhagen S

TABLE I
THE SYSTEMATIC RATING SYSTEM FOR THE MOAM MODEL.

	0 Points The mean is NOT CONSIDERED in any way.	**1 Point. The mean is implemented but has NO RELATION to the overall affective expression of the robot**	**2 Points. The mean provides a COORDINATED effort to increase the impact of the affective expression.**	**3 Points. The mean provides a RESPONSE to incoming stimuli.**
Morphology	The morphology has not been considered from an affective expression point of view	The morphology has been considered in the design of the robot, but is is unrelated to the affective expression.	The morphology is designed to support the overall affective reflection. Eg. colored lights express fear.	The morphology changes as a reaction to stimuli. Eg. adapts to context or reacts with lights.
Movement and Orientation	The robot does not move	The robot moves with no consideration to expessions	The robot moves to highlight emotional state	The robot moves as a response to stimuli.
Posture & Gestures	The robot has no onboard movement	The robot can move internally, but with no relation to conveying emotions.	The robot changes onboard positions or moves limbs to display emotional states.	The robot moves internally as a reaction to stimuli. Eg. it waves to other robots.
Sound	No sound or naturally occuring noise	Sound with no relation to affective expression. Eg. status messages	Coordinated sound to increase affective impact. Eg. music to set an athmosphere,	The sound is used as a mean to respond to incoming stimuli.
Anthropomorphic reflection	The robot bares no resemblance to any recognizeable character	The robot has a single or two features shared with a known character. E.g. has arms.	The robot resembles a known character and it supports an anthropomorphic interpretation	The robot's reaction to incoming stimul matches the expectation of the character.

downside to simplifying the model is that in a few cases the outcome could have similar profiles for robots that have very different real-life potentials.

A. *Expression modalities*

The inner parts of the model consist of the different modalities of the affective means. Through literature studies of papers from previous affective robot research projects, we have identified five categories of expression modalities. The five identified high-level modalities are "Morphology", "Movement and Orientation", "Posture and Gestures", "Sound", and "Anthropomorphic Reflection". Each part of the model corresponds to a modality and depicts the amount of effort directed towards these specific affective aspects of the robot. Further details on each of these categories will follow this overview. When depicting arbitrary robots using the model each of the categories are rated from 0 to 3, and the points are added to the corresponding section of the model to form a diagram over the different measures. When considering each part, the following criteria are used to establish a distribution of points:

- **0 Points:** The mean is **not present** in any capacity.
- **1 Point:** The mean is implemented but has **no relation** to the overall affective expression of the robot.
- **2 Points:** The mean is implemented and provides a **coordinated effort** to increase the impact of the overall affective expression.
- **3 Points:** The mean is implemented, increases the affective impact level and provides a **response** to incoming stimuli.

Table I displays the criteria for each of the identified categories. The criteria were selected from a combination of interaction theory and first-hand experience from affective robots. We emphasize the importance of responsiveness in the affective expression modalities as delays and lack of responsiveness tends to cause interaction outage [15]. It is important to stress, that the ratings of each category are neither an expression of positive or negative scores. A zero-rated robot on all accounts can be perfectly suited for certain tasks depending on the target context, and type of task it is designed to solve. This means that the role of the robot is important and can purposefully limit the amount of point distributed in specific categories. Eg. Paro the therapeutic seal robot has zero points distributed to the movement category because its main purpose is to stay still at the lap of the interacting user [16]. This is why any MOAM model of a robot is context-specific and comparing different models should be done in light of the role the robot fulfills. It is likewise important to mention that the ratings given to the robots in this survey are strictly interpreted from the information gathered in the referenced articles. This means that it has not been possible to obtain some of the details on specific mean categories. This may result in errors in the ratings stated for those categories.

B. *Responsiveness and adaption*

The final points in every category describe the level of responsiveness or adaptation the robot exhibits towards incoming stimuli. It depicts how well, in any of the categories, the robots respond to external context changes. E.g. the robot might make a sound every time it discovers another robot or human in the vicinity, or it might change position to orient itself towards any entities discovered in the working scenario. E.g. Limbu 2013 [4], enabled the 'CuDDler' robot to respond to audio stimuli with both gesture and sound.

A response might be immediate but could also manifest as a longer-lasting effort from the robot to dynamically fit its means of expression to match the recipient of the interaction or as an attempt to better align with a current working context. Examples of such are Miranda 2018 [17], where a robot alters its longer-lasting personal traits such as level of disagreeableness through an interaction.

C. Context and task alignment

The outer circle of the model corresponds to the working context as every individual working context demands different kinds of expressive means. E.g. low light situations makes gestures and postures hard to decipher while lights and sounds fit well. Even small changes to the context may demand large changes to the composition of expression methods to remain effective. This makes it difficult to create multi-purpose expressive means without dynamic ad-hoc adaption to the current environment and attention to the target of the interaction. As Bennett 2014 [18] argues, the context changes influence how we recognize the affective expressions of robots. If the context supports the expressed emotions the recognition rate will increase. Aligning the means of expression with the context can potentially amplify the conveyance of emotional values.

III. Using the MOAM model

The following example is provided to give an impression of the applicability of the model. A single robot has been selected and analyzed using the model in accordance with the criteria outlined in Table I. Stiehl et al. 2006 designed a therapeutic robot companion to function alongside nurses and improve the health and well being of the patients. It features several input sensors and reacts to touch, temperature, audio input, and visual stimuli. To express emotions the robot can change its posture, move limbs or emit sounds in response to user input. Furthermore, the robot has the appearance of a teddy bear, with fabric fur covering most of the body. The MOAM point distribution of that robot is depicted in the right image of Figure 2.

The affective scores in Figure 2 represent an affective robot with an even distribution of points. Although the robot is designed for the specific purpose of being stroked and to react to user input using movements and sound, the morphology has been considered in the design phase as well. The selected materials are applied to make the robot seem nice to touch and to highlight its familiar anthropomorphic shape and appearance of a teddy bear. The Huggable robot responds well to user input, and the responsive elements consisting of gestures, sound, and posture, increase the affective interpretation of its overall behavior.

The intended task for the huggable robot is to provide therapeutic comfort to medical patients, and as such, it is important that the robot's affective impression is considered. The 2 points in morphology mean that the robot's appearance and construction have been specially designed to support the affective impression. The Huggable has no locomotion

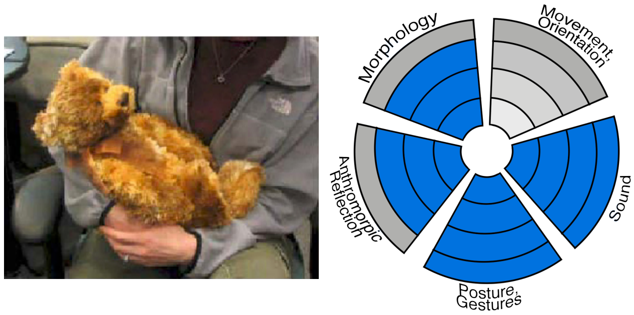

Fig. 2. Left: (Pending confirmation of image usage) The Huggable robot by Stiehl et al. 2006. Right: MOAM for Stiehl 2006 the huggable.

ability, resulting in a zero rating for the movement category. As the robot responds to outer stimuli with moving limbs and changing postures the robot is rated 3 for the onboard moving category. The same argument counts for the sound rating of 3. The robots resemble a well-known character type (a teddy bear), but the response manifestations to outer stimuli do not match the appearance of that figure, resulting in a rating of 2 for the anthropomorphic reflection.

IV. Modalities of measures

The following sections provide further details and illustrative examples for each category. All robots included in this survey have been rated using the MOAM model, and the resulting point distributions can be seen in Table II

A. Morphology

The morphology of a given robot describes its general appearance and depicts the level of attention given to affective expression details in the construction phase. A model with a high number of points allocated to the morphology category means, that the visual appearance or tactile sensation has been considered to a high degree from an affective expression perspective during the physical construction phase of the robot. This includes physical size and choice of colors [19], [20], type of materials used in the construction [3], [21], the sturdiness and build-quality or lack thereof [3], and the form and shape of the robot [21], [22]. Furthermore, as a mean of causing affective changes, the morphology precedes any initial contact and works at large spatial distances from the recipient, as long as there is a clear line of sight to the robot.

Some robot projects take advantage of factors that are already affiliated with certain types of signals. Eg. a red color means danger. Bethel 2009 used a blue light placed on the undercarriage of the robots to produce a calming effect [19]. Boccanfuso et al. 2015 used the 'Sphero' robot with colors, sound, and movement to simulate the expression of emotions [20].

The familiarity of certain appearances is also used by Singh et al. 2013 in the form of a dogtail attached to a small-sized robot [22]. Using the shape and size of recognizable animals was done by Sefidgar et al. 2016 in a small rat-like form factor therapeutic robot [21]. The soft materials were similarly important for the 'Huggable' robot introduced by

Stiehl et al. 2006 [3]. It made the humans that interacted with it relax when touching it and the fabric type added a teddy bear aesthetic to the robot.

B. Movement and orientation

The contents of this category are all implementations that influence how the robot moves, and how the robot reflects behavior of directing attention to something or someone in the vicinity. The specific speed [23], [24], acceleration changes [25], directional patterns [20], [24], [26], orientation [19], [27], and gait patterns employed by the robot as it moves from point A to B can convey emotional status. Yoshioka et al. 2015 and Boccanfuso et al. 2015 used simple small robots that employed movement style to successfully express emotions with changes in direction, velocity, acceleration and frequency of rotation [20], [24]. In Bethel et al. 2009 an emotive mode of the robots made them approach slowly, keep low to the ground and sustain an orientation towards the recipient to express attentiveness, caring and caution [19].

The relation between the acceleration curve and the type of interpreted delivered emotion was investigated in Saerbeck et al. 2010 [25]. The research results indicated a strong relation between motion parameters and affective recognition, such as causality between the acceleration curve and the PAD placement on arousal and valence axis. A system to retrofit existing robots was introduced in Fernandez et al. 2017 as an emotional enrichment system [26]. The aim was to enable users to describe emotions and to enrich the movements of the robot using these descriptions.

C. Posture and Gestures

As movements and orientation describe the positioning of the robot in an external or global scope, this category describe onboard or internal movements. This covers gestures [22], [28]–[36], speed of motions [32], [37], main body movements [4], [21], [38], posture [7], [28], [39], and touch [40]. With a combination of gestures, torso movement and facial expressions, Hegel et al. 2011 used a 'BarthocJr' model robot to measure and mimic live emotions from a human recipient [38]. The duplication of emotions and expressions mimicked by the robot works as a primitive form of empathy. Using solely body movements to express emotions, a faceless Nao robot in Cohen et al. 2011 successfully conveyed emotions as well as an 'iCat' robot that had a face [29], [41]. This indicated a high affective impact of solely employing body movements. Sefidgar et al. 2016 employed ears that stiffen in their therapeutic robot and a moving rib cage that simulated breathing [21]. Emotions can be expressed through touch as well. This was investigated in Chen et al. 2011, with a medical robot that touches its patients to calm them down [30]. The results indicated that the best effect was gained when people understood the intentions of the robot.

The effect of the neck, arm, and eyelid movement was researched by Limbu et al. 2013 in a study using the 'CuDDler' teddy bear therapeutic robot [4]. It was found that a combination of movements had a soothing effect on humans interacting with it. With a subsystem to generate emotions, Park et al. 2007 used a robot to show several emotion types including Fear, surprise, joy, anger, and sadness [36]. The robot employed motion in combination with posture and gestures and responded to user input from touch sensors whenever the users stroked it. A robotic stand-up comedian was developed in Addo et al. 2014 and it was discovered that using gestures enhanced the comedic impact on the audience [33]. Investigating affective physiology was the aim of Bianchi et al. 2016, with the development of an affective touch device built from rollers to simulate a pleasant human stroke [40]. The test persons could distinguish different kinds of emotional touch, which indicated that touch works as an effective way of expressing affect.

D. Sound

The sound aspect of an affective robot covers all audio originated from the robot. This includes both naturally occurring sounds (eg. the sound of wheels turning, limbs moving, servo buzzing), as well as artificial sounds emitted from the robot. The artificial sounds include voice [33], [42]–[45], soundscapes [46], [47], and notifications sounds [4], [20], [23], [48].

Matching the audio to the context is used by Lisetti et al. 2004 with a robot that has different voices to better match face and scenario [42]. Read and Balpaeme 2012 used non-linguistic sounds for robots to communicate with children and found that utterance rhythm is influential, while the pitch contour may have little importance in how the message is conveyed [48]. Gonsior et al. 2012 depicted emotions from the PAD space, by changing the voice with a different pitch, range, and accent [43].

To complement the behavioral traits of a pet dog robot, Yang et al. 2013 used audio as one of the expression modalities to convey both cognitive and emotional statuses [23]. Zhang et al. 2017 rated the importance of a robot's current synthetic emotional values with each other and formed the pitch, rate [47]. and volume of the robot's voice thereafter. The NAO robot platform was used by Winkle et al. 2017 to determine that the recognition of emotional values from robot voice and motion is possible in explicit validation experiments, but does not work with similar effect in socially assistive interaction situations [44]. The results suggest, that the correct interpretation of emotions relies on the human recipient to have formed an expectation of the attempted conveyed emotion. The impact of sound seems to work best when it matches the appearance of the robot, as was indicated by Becker et al. 2009 investigating laughter in robots [45]. The results were dependent on how well the synthetic laughter matched the robot appearance, and furthermore depended on the receiver's traits such as gender and nationality.

E. Anthropomorphic reflection

The anthropomorphic reflection attribute describes how much emphasis is placed on making the robot appear like a humanoid or recognizable character. Examples of humanoid inspired robots are Sophia [49], Gemini(s) [50], [51], Barthoc Jr. [52] and Pepper [53], while robots based on familiar characters include among others Leonardo [54], Aibo [55], and Paro [16]. There is currently an emphasis on using facial features in many robot research projects, [1], [2], [28], [42], [56]–[59], under the assumption that using a face makes it easier to convey emotions as a result of human familiarity with interpreting affective status through most social interactions. Only a small number of facial features are needed to successfully express emotions, as Bennet et al. 2013 found using only lips and eye lines to convey affective status [2].

Coupling a face with other means of expressions in a consistent manner over time could improve the amount of impact. This is a concept Lisetti et al. 2004 attempted to utilize in a service robot that maintained an ongoing personality throughout a series of interactions [42]. Zecca et al. 2008 designed KOBIAN a humanoid with the ability to convey emotions using facial expressions and by using bio-inspired body language [60]. It is not necessary to employ a whole face to trigger an emotional response, as Egawa et al. 2016 discovered using a single eye pupil in combination with an artificial laughter sound [61]. The results demonstrated that the dilated pupil response with a laughing response is effective for enhancing empathy.

V. Results and discussion

The table in Table II provides an overview of MOAM point distributions of the robots included in this paper. It is a table created based on information gathered through a literature study of papers published from previous affective robot projects. To create it, the authors gathered all information available from the paper regarding each specific category and distributed points according to the rules outlined in the matrix seen in Table I. By using the point distribution rules outlined in the MOAM matrix, the authors attempted to approach an objective overview of the affective means available for each robot. In the table, the robots are sorted by the number of non zero affective mean categories. 15% of the robots have points distributed to all categories while the remaining 85% display a single or several categories that contain a zero-rating. Furthermore, the average number of categories per robot with a zero-point distribution is 1.71. As most of the robots included in this paper are built for research purposes, this could indicate that it is the norm to focus solely on a single category when testing affective means. About a quarter of the included robots (25.6%) has a zero rating in more than 3 categories. It can be argued that disregarding several categories could have a negative impact on the robot's affective expression abilities. Eg. a research project might ask participants to consider the affective facial expressions of a robot, but forget to acknowledge the loud mechanical noise the robot is emitting throughout the tests. This could subconsciously influence how the robot is perceived by the audience.

The MOAM model illustrates the affective strengths of the robots, but the model also highlights the areas which

TABLE II

The distribution of points for robots included in this paper. The intensity of colors indicate the score.

Category / Publication	Morphology	Movement	Gestures	Sound	Anthropomorphic
Xu et al. 2015 using Nao	2	2	2	2	2
Zecca et al. 2008	2	2	2	2	2
Becker et al. 2009 Robovie-II	1	1	2	2	1
Fujita et al. 2001 (AIBO)	2	2	2	1	2
Becker et al. 2009 Robovie-R2	1	1	2	2	2
Pepper (Softbank robotics)	2	1	1	1	2
Singh et al. 2013	2	1	2	0	1
Yim et al. 2009	1	1	2	1	0
Stiehl et al. 2006	2	0	3	3	2
Yang et al. 2013	1	2	2	2	0
Park et al. 2007 CuDDler	2	0	3	2	3
Cohen et al. 2011 w. NAO	2	2	2	0	2
Chen et al. 2011	1	0	2	2	1
Xu et al. 2013	2	1	2	0	2
Sophia (Hansen Robotics)	2	0	3	3	3
Gemini (By Hi. Ishiguro)	2	0	2	3	3
Breazeal et al. 2004	3	0	3	1	3
Paro (By Parorobots)	2	0	1	1	2
Breazeal et al. 2003 Kismet	1	0	3	1	1
Breemen et al. 2005	3	0	2	0	2
Hegel et al. 2006	2	0	2	0	3
Bethel et al. 2009	2	2	0	2	0
Boccanfuso et al. 2015	2	2	1	0	0
Sefidgar et al. 2016	2	0	3	0	3
Park et al. 2007	1	2	3	0	0
Fernandez et al. 2017	1	2	0	0	2
Addo et al. 2014 Zoei	2	0	2	0	2
Lisetti et al. 2004 Cherry	1	1	0	1	0
Gonsior et al. 2012	0	1	0	1	2
Yoshioka et al. 2015	1	2	0	0	0
Zhang et al. 2017	0	0	0	3	0
Benson et al. 2016	1	0	0	0	2
Egawa et al. 2016	0	0	1	0	1
Saerbeck et al. 2010	0	2	0	0	0
Fernandez et al. 2017	0	2	0	0	0
Yoshioka et al. 2015	0	2	0	0	0
Lee et al. 2013	0	0	2	0	0
Bianchi et al. 2016	0	0	2	0	0
Bennett et al. 2013	0	0	0	0	1

represent opportunities for improvements. A fully covered inner circle of the model equals a distribution of two points for each category of means. This requires each category to be manifested in some form but demands no further coordinated effort to increase the affective impact of the robot. It is possible, that by ensuring a point distribution that covers the inner circle when constructing robots, the outcome could be more efficient affective robots with lesser disregarded areas to influence how they are perceived. Even some state-of-the-art affective robots contain categories of the MOAM model with zero points allocated leaving room for further improvements. An unattended category of means could present an opportunity to add further expression means to mitigate any negative aspects of the category.

The lower entries of the table consist of robots that are constructed with an emphasis on a single category of means. This makes a lot of sense as these robots are often designed to test the validity of a single mean of affective expression. There may be practical (and economic) reasons for limiting the number of included affective design details. However, the indication that the MOAM categories could influence each other in both positive and negative ways, could be viewed as an argument for considering other means of expression when designing affective systems. The distribution of points in each MOAM model can in some cases be limited by the task intended for the robot to handle (Eg. Paro [16] which it not designed to move), making it difficult to compare robots intended for different contexts. For that reason, there might be a research opportunity in exploring how to create further specialized MOAM models containing the attributes of specific working scenarios. Such models could provide an easier method to compare robots designed to fulfil similar roles (Eg. social companion robots, therapeutic robots, robot teachers). However, the main intention of proposing the model is to provide a general overview of the technical capabilities of each robot, not to rate how well the robot performs in different working contexts. As such, the current model reflects a loss of finer details to gain a wider range of included robots to compare with. Furthermore, the model is not solely intended to provide a scoring mechanism. It has a purpose besides working as a comparison between robots, it is also intended as a quick reference to aid in the process of designing affective robots. The overview table in this paper was created solely by the authors. To generalize the result, future iterations should also include participation from a larger sentiment of people to minimize the influence of subjective evaluation. The aim of using the point distribution matrix was to avoid bias and subjectivity in the creation of MOAM models. However, some categories are less prone to subjectivity than others (Eg. the anthropomorphic reflection can be culturally dependent), but this can be mitigated by evaluating the robot in light of its intended role and working context. Doing so increases the consistency of the resulting MOAM points distributed by different people.

Some research projects build upon commercially available robots such as the NAO robot when testing affective means. Building on top of these platforms allows the research teams to emphasize on different aspects of affective means. Using a common base for affective robot research is a good idea as it works towards minimizing any negative impact from disregarded categories. However, relying on NAO and similar robot solutions could in some situations mean missing an opportunity to customize the morphology to the specific context. It is possible that some scenarios could demand an easier customizable affective robot to better align with the working context but that could be a possible topic for further research.

VI. CONCLUSION

To this date, essential progress has been made in affective robot research. As a result, we have substantial knowledge of how single means of expression works. In comparison, we know relatively little about how categories of means influence each other when used together or when disregarded.

The paper has identified five high abstraction level categories of expressional means and has provided examples of each to highlight their functionality. The mean categories are 'Morphology', 'Posture and Gestures', 'Sound', 'Movement' and 'Anthropomorphic reflection'. All identified categories have been summed up in a proposed model of affective means (MOAM) to capture strength and weaknesses for any robot from an affective perspective. To make the MOAM models comparable to each other, this paper has also proposed a point distribution system to allocate points to the affective mean categories. The MOAM model and its underlying point distribution system have been used to score and compare all included affective robots. The resulting MOAM scores are directly comparable and work as systematic descriptors of the affective strengths and shortcomings of the robot, but should be viewed in light of the intended role and working context of each robot. Overall we argue, that the MOAM model can sufficiently categorize and be used to compare a large plethora of different affective robot types.

Using the models to rate robots and compare with each other gave insight to the possible opportunities for improving even the already successful robots that rely on single affective means to express affective states. There are indications that the synergies between affective means could possibly change the impact of the overall impression of the robot. These indications warrant further investigation into the feasibility of testing single means of affection in isolation. Furthermore, we argue that the number of robots in this survey, that has room to add further means of expression, shows there could be an unexplored area of research in building more complete affective agents with attention to all categories.

REFERENCES

[1] C. Breazeal, "Emotion and sociable humanoid robots," *Int. J. Hum.-Comput. Stud.*, vol. 59, pp. 119–155, 2003.

[2] C. C. Bennett and S. Sabanovic, "Perceptions of affective expression in a minimalist robotic face," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 81–82, 2013.

[3] W. D. Stiehl, C. Breazeal, K.-H. Han, J. Lieberman, L. Lalla, A. Z. Maymin, J. Salinas, D. Fuentes, R. L. Toscano, C. H. Tong, A. Kishore, M. Berlin, and J. Gray, "The huggable: a therapeutic robotic companion for relational, affective touch," *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.*, vol. 2, pp. 1290–1291, 2006.

[4] D. K. Limbu, C. Y. A. Wong, A. H. J. Tay, T. A. Dung, Y. K. Tan, T. H. Dat, A. H. Y. Wong, W. Z. T. Ng, R. Jiang, and L. Jun, "Affective social interaction with cuddler robot," *2013 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, 2013.

[5] A. Mehrabian, *Basic dimensions for a general psychological theory : implications for personality, social, environmental, and developmental studies*. Cambridge : Oelgeschlager, Gunn and Hain, 1980.

[6] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 305–311, 2011.

[7] M. Rehm, "Experimental designs for cross-cultural interactions: A case study on affective body movements for hri," *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pp. 78–83, 2012.

[8] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[9] P. R. D. Silva, A. P. Madurapperuma, S. G. Lambacher, and M. Osano, "Therapeutic tool for develop child nonverbal communication skills through interactive game," *International Conference on Computational Inteligence for Modelling Control and Automation*, pp. 33–33, 2006.

[10] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vision Comput.*, vol. 27, pp. 803–816, 2009.

[11] J.-W. Tan, S. Walter, A. Scheck, D. Hrabal, H. Hoffmann, H. Kessler, and H. C. Traue, "Facial electromyography (femg) activities in response to affective visual stimulation," *2011 IEEE Workshop on Affective Computational Intelligence (WACI)*, pp. 1–5, 2011.

[12] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, pp. 311–324, 2007.

[13] S. McGilloway, R. Cowie, C. ED, S. Gielen, M. Westerdijk, and S. Stroeve, "Approaching automatic recognition of emotion from voice: A rough benchmark," *Proceedings of the ISCA Workshop on Speech and Emotion*, 01 2000.

[14] S. Koelsch, "Brain correlates of music-evoked emotions," *Nature Reviews Neuroscience*, vol. 15, pp. 170–180, 2014.

[15] C. Kohrs, N. Angenstein, and A. Brechmann, "Delays in human-computer interaction and their effects on brain activity," *PLOS ONE*, vol. 11, p. e0146250, 01 2016.

[16] T. Shibata, K. Inoue, and R. Irie, "Emotional robot for intelligent system-artificial emotional creature project," in *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, Nov 1996, pp. 466–471.

[17] J. Martínez-Miranda, H. Espinosa, I. Espinosa Curiel, H. Avila-George, and J. Rodríguez-Jacobo, "Age-based differences in preferences and affective reactions towards a robot's personality during interaction," *Computers in Human Behavior*, 03 2018.

[18] C. C. Bennett, S. Sabanovic, M. R. Fraune, and K. Shaw, "Context congruency and robotic facial expressions: Do effects on human perceptions vary across culture?" *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pp. 465–470, 2014.

[19] C. L. Bethel, K. Salomon, and R. R. Murphy, "Preliminary results: Humans find emotive non-anthropomorphic robots more calming," *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 291–292, 2009.

[20] L. Boccanfuso, E. S. Kim, J. C. Snider, Q. Wang, C. A. Wall, L. DiNicola, G. Greco, F. Shic, B. Scassellati, L. Flink, S. Lansiquot, K. Chawarska, and P. Ventola, "Autonomously detecting interaction with an affective robot to explore connection to developmental ability," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7, 2015.

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

[22] A. Singh and J. E. Young, "A dog tail for communicating robotic states," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 417–417, 2013.

[23] J.-Y. Yang, B. K. Dinh, H.-G. Kim, and D.-S. Kwon, "Development of emotional attachment on a cleaning robot for the long-term interactive affective companion," *2013 IEEE RO-MAN*, 2013.

[24] G. Yoshioka, T. Sakamoto, and Y. Takeuchi, "Inferring affective states from observation of a robot's simple movements," *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 185–190, 2015.

[25] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 53–60, 2010.

[26] J. M. A. Fernandez and A. Bonarini, "Enriching robot's actions with affective movements," *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 424–429, 2017.

[27] C.-Y. Lin, L.-C. Cheng, C.-C. Huang, L.-W. Chuang, W.-C. Teng, C.-H. Kuo, H.-Y. Gu, K.-L. Chung, and C.-S. Fah, "Versatile humanoid robots for theatrical performances," *International Journal of Advanced Robotic Systems*, vol. 10, p. 1, 01 2013.

[28] J.-D. Yim and C. Shaw, "Designing cally, : a cell-phone robot," in *CHI Extended Abstracts*, 2009.

[29] I. Cohen, R. Looije, and M. A. Neerincx, "Child's recognition of emotions in robot's face and body," *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 123–124, 2011.

[30] T. L. Chen, C.-H. King, A. L. Thomaz, and C. C. Kemp, "Touched by a robot: An investigation of subjective responses to robot-initiated touch," *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 457–464, 2011.

[31] J. Xu, J. Broekens, K. V. Hindriks, and M. A. Neerincx, "Effects of a robotic storyteller's moody gestures on storytelling perception," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 449–455, 2015.

[32] ——, "The relative importance and interrelations between behavior parameters for robots' mood expression," *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013.

[33] I. D. Addo and S. I. Ahamed, "Applying affective feedback to reinforcement learning in zoei, a comic humanoid robot," *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pp. 423–428, 2014.

[34] W. Johal, S. Pesty, and G. Calvary, "Toward companion robots behaving with style," *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, vol. 2014, pp. 1063–1068, 08 2014.

[35] L. Rincon, E. Coronado, H. Hendra, J. Phan, Z. Zainalkefli, and G. Venture, "Expressive states with a robot arm using adaptive fuzzy and robust predictive controllers," *2018 3rd International Conference on Control and Robotics Engineering (ICCRE)*, pp. 11–15, 2018.

[36] C. Park, J. Ryu, J. Sohn, and H. Cho, "An emotion expression system for the emotional robot," *2007 IEEE International Symposium on Consumer Electronics*, pp. 1–6, 2007.

[37] M. Memarian, R. B. Gorbet, and D. Kulic, "Control of soft pneumatic finger-like actuators for affective motion generation," *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1691–1697, 2015.

[38] F. Hegel, T. Spexard, B. Wrede, G. Horstmann, and T. Vogt, "Playing a different imitation game: Interaction with an empathic android robot," *2006 6th IEEE-RAS International Conference on Humanoid Robots*, pp. 56–61, 2006.

[39] H. Lee, Y. Kim, and M. Kim, "Come on in!: A strategic way to intend approachability to a space by using motions of a robotic partition," *2013 IEEE RO-MAN*, pp. 441–446, 2013.

[40] M. Bianchi, G. Valenza, A. Greco, M. Nardelli, E. Battaglia, A. Bicchi, and E. P. Scilingo, "Towards a novel generation of haptic and robotic interfaces: Integrating affective physiology in human-robot interaction," *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 125–131, 2016.

[41] A. J. N. van Breemen, X. Yan, and B. Meerbeek, "icat: an animated user-interface robot with personality," in *AAMAS*, 2005.

[42] C. L. Lisetti, S. M. Brown, K. Alvarez, and A. H. Marpaung, "A social informatics approach to human-robot interaction with a service social robot," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, 2004.

[43] B. Gonsior, S. Sosnowski, M. Buß, D. Wollherr, and K. Kühnlenz, "An emotional adaption approach to increase helpfulness towards a robot,"

2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2429–2436, 2012.

[44] K. Winkle and P. Bremner, “Investigating the real world impact of emotion portrayal through robot voice and motion,” *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 627–634, 2017.

[45] C. Becker-Asano and H. Ishiguro, “Laughter in social robotics–no laughing matter,” *Intl. Workshop on Social Intelligence Design (SID2009)*, 01 2009.

[46] J. H. Lui, H. Samani, and K. yu Tien, “An affective mood booster robot based on emotional processing unit,” *2017 International Automatic Control Conference (CACS)*, pp. 1–6, 2017.

[47] R. Zhang, Y. He, and Z. Liu, “Communication atmosphere in humans and robots interaction based on fuzzy analytical hierarchy process,” *2017 36th Chinese Control Conference (CCC)*, pp. 6767–6772, 2017.

[48] R. Read and T. Belpaeme, “How to use non-linguistic utterances to convey emotion in child-robot interaction,” *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 219–220, 2012.

[49] D. Hanson, *Humanizing Robots: How making humanoids can make us more human*. Amazon Digital Services LLC, 2017.

[50] S. Nishio, H. Nakanishi, and T. Fujinami, Eds., *Investigating Human Nature and Communication through Robots*. Frontiers Media, Jan. 2017.

[51] C. W. Becker-Asano and H. Ishiguro, “Evaluating facial displays of emotion for the android robot geminoid f,” *2011 IEEE Workshop on Affective Computational Intelligence (WACI)*, pp. 1–8, 2011.

[52] F. Hegel, S. Gieselmann, A. Peters, P. Holthaus, and B. Wrede, “Towards a typology of meaningful signals and cues in social robotics,” *2011 RO-MAN*, pp. 72–78, 2011.

[53] SoftbankRobotics, “Pepper,” nov 2018.

[54] C. Breazeal, A. G. Brooks, J. Gray, G. Hoffman, C. D. Kidd, H. Lee, J. Lieberman, A. L. Thomaz, and D. Chilongo, “Tutelage and collaboration for humanoid robots,” *I. J. Humanoid Robotics*, vol. 1, pp. 315–348, 2004.

[55] M. Fujita, “Aibo: Toward the era of digital creatures,” *I. J. Robotics Res.*, vol. 20, pp. 781–794, 2001.

[56] D. Benson, M. M. Khan, T. Tan, and T. Hargreaves, “Modeling and verification of facial expression display mechanism for developing a sociable robot face,” *2016 International Conference on Advanced Robotics and Mechatronics (ICARM)*, pp. 76–81, 2016.

[57] G.-Y. Park, S.-I. Lee, W.-Y. Kwon, and J.-B. Kim, “Neurocognitive affective system for an emotive robot,” *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2595–2600, 2006.

[58] R. Gockley, R. G. Simmons, and J. Forlizzi, “Modeling affect in socially interactive robots,” *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 558–563, 2006.

[59] S.-I. Lee, G.-Y. Park, and J.-B. Kim, “Natural emotion expression of a robot based on reinforcer intensity and contingency,” *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 2144–2149, 2007.

[60] M. Zecca, N. Endo, S. Momoki, K. Itoh, and A. Takanishi, “Design of the humanoid robot kobian - preliminary analysis of facial and whole body emotion expression capabilities-,” *Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots*, pp. 487–492, 2008.

[61] S. Egawa, Y. Sejima, Y. Sato, and T. Watanabe, “A laughing-driven pupil response system for inducing empathy,” *2016 IEEE/SICE International Symposium on System Integration (SII)*, pp. 520–525, 2016.

12

Paper 2: On the causality between affective impact and coordinated human-robot reactions

On the causality between affective impact and coordinated human-robot reactions

Morten Roed Frederiksen[1] Kasper Stoy[2]

***Abstract*— In an effort to improve how robots function in social contexts, this paper investigates if a robot that actively shares a reaction to an event with a human alters how the human perceives the robot's affective impact. To verify this, we created two different test setups. One to highlight and isolate the reaction element of affective robot expressions, and one to investigate the effects of applying specific timing delays to a robot reacting to a physical encounter with a human. The first test was conducted with two different groups (n=84) of human observers, a test group and a control group both interacting with the robot. The second test was performed with 110 participants using increasingly longer reaction delays for the robot with every ten participants. The results show a statistically significant change ($p<.05$) in perceived affective impact for the robots when they react to an event shared with a human observer rather than reacting at random. The result also shows for shared physical interaction, the near-human reaction times from the robot are most appropriate for the scenario. The paper concludes that a delay time around 200ms may render the biggest impact on human observers for small-sized non-humanoid robots. It further concludes that a slightly shorter reaction time around 100ms is most effective when the goal is to make the human observers feel they made the biggest impact on the robot.**

I. Introduction

Creating robots that can understand and express emotions is a many-faceted problem. One of the many challenges lies in designing a relatable robotic behavior with which people will want to interact. If we disregard digital communication channels, robots convey information through simple means of expression that includes: Sound, appearance, movements, and gestures [1]. These means can improve how well the intentions of the robot are understood, and correctly timing when to use them can further improve the interaction and can influence how the robot is perceived [2]. A lot of research has focused on the expressive abilities of robots and have so far accomplished making people recognize robotic expressions of emotions using morphological attributes [3]–[5], facial features [6]–[13], movement [14]–[16], orientation [17], [18] sound [8], [19]–[24], and gestures [5], [13], [20], [25]–[32]. When it comes to expressing affective information and standard emotions, many projects focus on how to maximize comprehension. Relatively few projects in comparison focus on the impact of delaying when the expressive features of the robot are used, and how the causality between participant and robot reactions can affect how the affective information is conveyed. Michael 2010 proposes how perceived shared emotions can facilitate coordination between interacting humans without either of them possessing previous knowledge of intentions [33]. This paper focuses on whether this effect is equally present in human-robot-interactions and investigates the following:

- If there is a causality between reaction coordination and perceived affective impact on a robot. In other words: When humans and robots react to the same event, will the humans perceive the robots' reactions as stronger?
- Whether delaying the reactions of a robot in a physical conflict interaction can strengthen its perceived affective impact.

Gaining knowledge on these aspects of expression abilities is something all areas of robotics can benefit from. The investigation may provide an answer to when and how robots should behave in order to strengthen the affective impact of an interaction. This could be beneficial in situations where robots are required to convey vital information as efficiently as possible. E.g. Socially assistive robotics and rescue robots that operate in demanding working environments may be vastly improved if we, by altering how and when they use their communicative features, can make them communicate better in a critical situation.

Through each human-robot interaction, the timing dictates who initiates actions throughout the encounter. E.g. a swift reacting robot could make a human recipient hold back in the interaction or a robot that delays answering could make a human counterpart take charge of the situation. Among other aspects of communication, the timing encompasses both estimating when to perform movements (for robots to safely cooperate with humans) and controlling the flow of dialogue between humans and robots [34], [35].

When robots react to something, the reaction highlights the

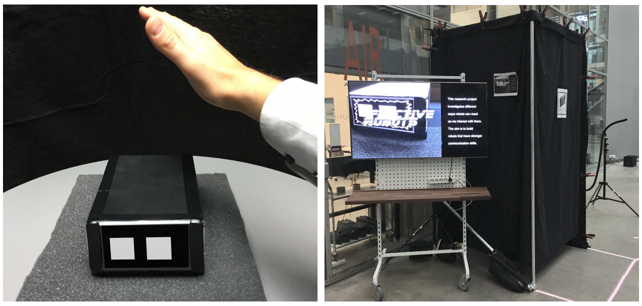

Fig. 1. Left: The "Affecta" robot. The robot was fastened to a soft foam pad to hinder it from moving as people interacted with it. Right: The test setup included an isolated room to let the participants interact with the robot undisturbed.

[1]Morten Roed Frederiksen mrof@itu.dk and [2]Kasper Stoy ksty@itu.dk are affiliated with the REAL lab at the Computer science department of The IT-University of Copenhagen, Rued Langgaards vej 7, 2300 Copenhagen S

connection between robot behavior and the context event, and it establishes the direction for the current communication. E.g. for a robot that is designed to portray being afraid of a dog in the vicinity, there is a timed frame of opportunity after the dog initiates an action where the robot can react. Any reactions applied in connection with the dog's actions will be perceived as connected to that event or that agent in the scenario. The robot's reaction will be interpreted in light of the event and if a human experiences the same event, the shared experience may be used to establish a connection between the human and the robot. The reaction time and response in the situation is influenced by the complexity and familiarity of the event information as outlined in Hyman 1953 [36]. Besides the complexity of event information to which the robot reacts, the hypothesis is that the following two things (among other factors) can influence how the expression of a robot's reaction is perceived:

- The time delay with which the robot reacts
- If the reaction is shared with someone.

To investigate this, we used two experiments. The first test was a standard A-B test aimed at isolating the effects of coordinating human-robot reactions to a context event, while the second test focused on how reaction delays affected the shared experience in a physical interaction. Our findings show a causality between human-robot reaction coordination and the perceived arousal level of the test robots, with a statistically significant ($p<.05$) difference between the main group and the control group. The results further indicate that the reaction times of the robots in physical interactions influence the affective state of the humans interacting with it. We argue that near human-like reaction reflexes overall have the biggest affective impact on the test participants, while a slightly lesser delay time (˜100ms faster) should be used when the aim is for the test participants to feel they made a big impression on the robot. The results also indicate that the perceived affective impact of the robot is strengthened slightly by delaying the reaction.

The presented findings are novel in that they present a new context for using shared experiences to gain emotional coordination in human-robot interaction scenarios. The new approaches are based on using non-humanoid robots and by placing participants on the opposite side of the robot in a high-intensity conflict situation. The results introduce many opportunities for further research on the topic. As a whole, they suggest investigating to what extend shared reactions could strengthen the affective expression abilities of rescue robots and improve the reception of critical messages in high-intensity contexts.

II. Other approaches

The timing aspects of cooperative interaction was investigated by Pan et al. 2019 [37] by in- and decreasing the reaction times of a robot that was handed an object. The study, which used a humanoid torso robot with a head and arms, found that the people preferred reaction time equal to normal human reaction time when interacting with the robot. Their test scenario was different than the scenario investigated in this paper, as it contained a low-intensity interaction, a humanoid robot, and a cooperative task to accomplish in the tests, whereas this project focuses on non-humanoid robots in a high-intensity scenario and a test task that emphasizes the conflict between the interacting human and robot participants.

Previous robot projects have investigated increasing the understanding of affective communication in their research. Brazeal et al. 2003 employed an emotional subsystem for the robot Leonardo and controlled realistic employment of several affective means of expression making it easier to understand [10]. Gunes et al. 2011 used a LEGO-based custom robot to convey the emotional intentions of classical music. The robot employed several affective means of expression including movement and onboard gestures to communicate the affective status [38]. The timing aspects were the focus of Huber et al. 2008 in which they investigated different ways of letting robots hand over objects to humans. Successfully handing over the objects requires both parties of the interaction to agree on a common timing for the involved movements. The study found that the less jerky the movement was, the safer they felt around the robot. [39].

Bing and Michael 2012 investigated how sharing a stressful experience with a humanoid robot can potentially help humans overcome the uncanny valley effect [40], [41]. The 2012 paper found that their test participants preferred familiar humanoids with whom they had shared a stressful experience with rather than familiar robots that they had shared a pleasant experience with. This paper aims to extend the results found in that paper on two different levels. It investigates whether the results are similar for a non-humanoid robot that bears no resemblance to a person, and it attempts to discover whether the result is isolated to people that are on opposite sides of a conflict- and stressful situation. This paper emphasizes how humans perceive robot specific nonverbal behavior which is also the focus in Putten et al. 2018 [42]. In this paper, the robot-like specific behavior is found less effective than using human-like familiar behaviors to convey affective information. Both Bing & Michael 2012 and Ptten et al. 2018 indicate the strengths of using human-inspired behaviors and morphology in their studies which make a good contrast to the experiments performed in this paper using non-humanoids and strictly robot-specific behaviors.

III. Method

The first test aims at investigating changes to the general composition of emotions, while the second test expands the investigation into a physical and confrontational scenario to see how that influences the perceived intentions of a robot. The second test also focuses on the immediate delay between the context event and the subsequent robot reaction to see how delaying the robot's reaction influences how the robot was perceived. As stated in Bing et al. 2012, a shared stressful event works stronger using humanoid robots, which is why a conflicting scenario with a non-humanoid robot was interesting for the second test in this paper [40].

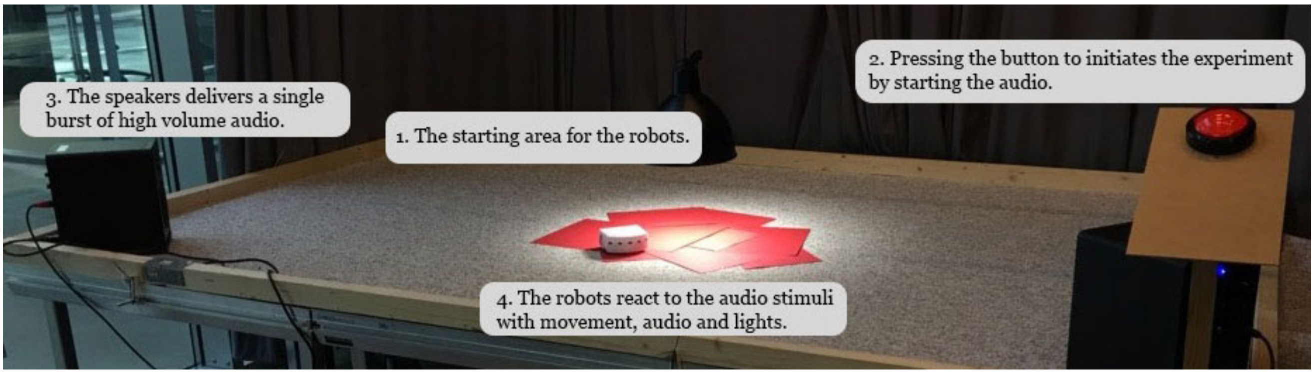

Fig. 2. The test setup for the initial experiment. The red center of the arena marks the starting position for the robots. The red button on the right side initiates the experiment in the first test. The same button was removed for the control test of the experiment in which the robots reacted with random intervals.

A. Using standard descriptors

In affective robotics research it is often the Pleasure, Arousal and Dominance (or PAD) scale that is used to describe emotional states [43], [44], while temporal aspects can be classified in the Traits, Attitudes, Moods and Emotions (TAME) architecture [45].

We quantify the affective impact by measuring the changes to the robot's perceived current emotion in PAD space. We measure differences between the two test groups on how the robot's affective state is perceived. If the test participants find it more or less pleasant, aroused, or dominant. E.g. if a person is angry during an interaction with a robot, and the robot emits a soothing sound to make the person change to a happier state, the angry emotion could move along the 'arousal' axis towards less aroused - which would be considered an affective change to the current affective state. This is what we use as a quantitative measure for the effects of coordinated reactions in the initial tests.

The tests followed a standard A-B pattern with two individual groups of test participants where one of them acted as a control group. The two groups would encounter the same scenarios, but the control group of participants would not experience coordinated reactions with the robots as they would react at random and out of phase with the participants. The test setup is depicted in Figure 2.

B. Moving to a physical interaction

Building upon the outcome of the first tests, the second test focused on how the shared reaction was perceived when the interaction context was changed to a physical and conflicting encounter with closer proximity between the participants and the robot. In this test, we asked the participants to physically strike the robot as much as they wanted and observe the reaction. We departed from using the standard PAD descriptor as we were not focusing on the composition of the affective impact, but rather on investigating where the interaction was perceived as making the largest impact - on the robot itself or the test participants. We also wanted to see how the delay time influenced the perceived size of the affective reaction and introduced delays between the physical interaction and the robot's reaction to highlight the connection between them. As the robot reacted in this context, the swiftness of the reaction made it more similar to a reflex than a prepared response. This approach was chosen as it matched the conflicting scenario. The sharing in the second test was solely the interaction, and we attempted to investigate how placing the participants and the robot on opposite sides of a conflict situation influenced the human-robot relationship.

IV. Experimental setup

In the first test, there were two groups with 42 people observing the robots in each of them. The overall gender distribution was 39 females and 45 males in ages from 10 to 50+. The majority of the participants were between 20 and 30 (71%) years old, and most of the participants either worked - or studied at The IT-University of Copenhagen (82.5%).

In the second test there were 110 participants distributed in 7 groups. The gender distribution here was 56% male and 44% female and the largest age group was 20-30 years old (33%) followed by people between 10-20 (20.2%). The initial test used a "Thymio 2" robot while the second test used and altered a custom-built "Affecta" robot designed to convey affective information.

A. The first test: impact of reaction

The setup of this test was comprised of three "Thymio 2" robots and a designated arena for the robots to move on. The arena was constructed from stage parts, forming a 220cm times 300cm surface, with floor carpets on top to create a smooth surface to easily maneuver on for the low-clearance Thymio 2 robots. The edges of the designated test arena were padded with a small wooden edge to prevent the robots from falling to the ground. The edges were fastened just high enough to trigger the proximity sensors positioned at the front, side, and back of the robots. The first test contained two experimental phases with different groups participating in each experiment. The tests were initiated in isolation from each other and followed this test outline:

Test steps:

1. The robots were initially placed at the center of the

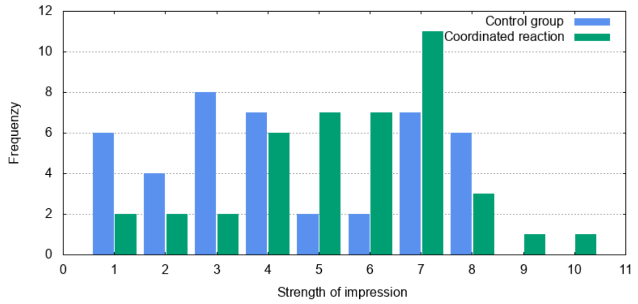

Fig. 3. The diagram shows the perceived arousal level of the robots. The blue-colored values are from the control test with uncoordinated human-robot reactions while the red-colored values are from the test group coordinated reaction between the participants and the robots.

arena. See Figure 2 for the initial position of the robots. 2. The participants were asked to start the experiment by pushing a button. 3. A high volume sound of an explosion was played as the button was pushed and the robots (and participants) reacted to the sound displaying fear. The robots used the following expression modalities: Sound, movement, and colored lights to convey the fear behavior. 4. The robots moved from the start area with maximum speed while displaying lights, and playing alerting audio signals in an attempt to show fearful behavior. 5. After 2 seconds of employing audio and lights, the robots continued to move but the audio and lights were turned off. This was done to enforce the connection between the reaction and the event that initiated it. 5. The robots moved randomly around on the surface while using front and back sensors to avoid the perimeter. 6. Once the robots encountered the center 'resting' area again, they stopped and waited until reacting again (start over from point 2).

The control group would go through the same steps. However, the robots would not react in coordination with the sound but at random intervals. After each experiment, the test participants were asked how *aroused* they perceived the robots were, how *pleasant* they perceived the robots found the experiment, and how *dominant* they perceived the current emotion for the robots was on a scale from 1 to 10 (1 meaning: not at all and 10 meaning: maximum possible). The participants were additionally asked to state their gender, and age.

B. The second test: the impact of specific timing

The second test used a custom-built robot as depicted in the left image of Figure 1. The robot was a small non-humanoid box-shaped robot that was designed to have implementations for a large variety of expression modalities, making it a great fit for this project. This specific robot design was 3d-printable, and suited the test setup. For the robot to remain stable for the physical interaction, only the top part of the robot was used and the bottom drive wheels not added. The robot consisted of two separate software architectures - a ROS based part to control the physical movement and gestures of the robot and a mobile application with access to all available sensors on a mobile smartphone. For this test, the mobile IOS based platform was expanded with a module for detecting physical movement using the onboard accelerometer. When the user would hit the robot the accelerometer sensor was triggered which informs the main robot controller to display a reaction using the mobile phone screen and audio capabilities of the robot (also supplied by the phone). The reaction consisted of a loud alert noise and jagged lines flashing at the edge of the screen. The second test was set up in a specially constructed and isolated test booth. The booth, which can be seen in the right image of Figure 1, contained a table with the robot at a raised position to facilitate a close proximity interaction, and it contained a poster with instructions for the test participants to strike the robot. One at a time we asked them to enter the test booth and hit the robot as much as they liked. They would interact with the robot by hitting it and observe how the robot reacted. When the test participants were finished with the physical interaction, they would step outside of the test booth and we proceeded by asking the following questions:

- How big an impact did your actions make on the robot?
- How big an impact did the robot's reaction make on you?
- How appropriate would you rate the robot's actions as being in light of how you interacted with it?

The participants were also asked to state their age group and their gender. The test was completed with 110 test participants. With each group of ten participants, the reaction delay of the robot's reactions was doubled starting from an initial reaction delay of 50ms ending at a reaction delay of 3200ms.

V. Results

The first test isolated the effects of coordinating human-robot reactions to a context event, while the second test used increasingly longer reaction delays to investigate how that affected the perceived affective impact of a human-robot physical interaction.

The results show three important findings:

- There is a causality between coordinating the reactions of humans and robots and the perceived arousal level of the test robots.
- The reaction times of the robots in physical interactions influence the affective state of the humans interacting with it and near human-like reaction reflexes (~250ms) have the biggest affective impact on the test participants
- A slightly lesser delay time (~100ms faster) is preferred when the aim is for the test participants to feel they made a big impression on the robot.

A. The influence of coordinating human-robot reactions

In the first test, we asked the participants to rate how *aroused* the robots seemed, and the difference between levels of perceived arousal was statically significant (Two-tail Wilcoxon signed-rank, $p<.05$). This shows a strong connection between experiencing a shared reaction with the robots and the interpreted level of arousal conveyed by the robots. The distribution of answers for the question on the perceived level of arousal can be seen in Figure 3, and the key figures for the same question can be seen in Table I.

We also asked the participants to rate the perceived pleasantness of the experience for the robots. The results for that question showed no relevant differences between the random group and the reaction group. The participants agreed that the experience was mildly unpleasant for the robots in both groups with key figures as seen in Table I. The last question regarded the perceived level of dominance for the current emotion, on which the participants rated each group with near similar scores. This indicated that there was no connection between the dominance level and sharing a reaction or not.

B. Reaction delays strengthen affective impact

In the second test, the results indicate that there was a preferred reaction delay around 200ms for the question regarding the perceived impact of the robot's actions on the participants who interacted with it. The resulting averages for that question can be seen in Figure 4. This enforced the results found in by Pan et al. 2019 and extends the finding to also include non-humanoid robots and a conflicting scenario rather than a cooperative context [37]. The results show that the robot made the biggest affective impact on the participants when it reacted to the physical interaction with human-like reaction times (which we assume is approximately 250ms). It is important to state that although our number of participants is relatively high (n=91), using the arithmetic mean for smaller individual groupings could make the result more easily affected by outliers.

We asked the participants to rate how big an impression the test participants' actions made on the robot, and for that question, the relative highest rated delay time was 100ms. This and the previous result indicate the following:

- If the aim is for the robot to make a big impression on the participants, it should react with near-human reaction times.
- If the aim is for the test participants to feel they made a big impression on the robot, it could benefit from reacting with a slightly smaller delay. (˜100ms faster).

We also asked the participants to rate the appropriateness of the robot's action in relation to the actions performed by the test participants. The resulting ratings were near at par with each other with a reaction time of 100ms rated relatively

	No Reaction (avg/dev)	With Reaction
Agitatedness	4.40/2.42	5.55/2.04
Pleasantness	4.83/2.27	4.71/2.11
dominance	3.76/2.34	3.98/2.50

TABLE I

THE AVERAGES AND STANDARD DEVIATION FOR THE ANSWERS FOR THE PERCEIVED LEVEL OF AROUSAL, PLEASANTNESS, AND LEVEL OF DOMINANCE IN THE TESTS WHERE THE PARTICIPANTS SHARED A REACTION WITH THE ROBOTS AND THE CONTROL TEST IN WHICH THE ROBOTS REACTED AT RANDOM INTERVALS.

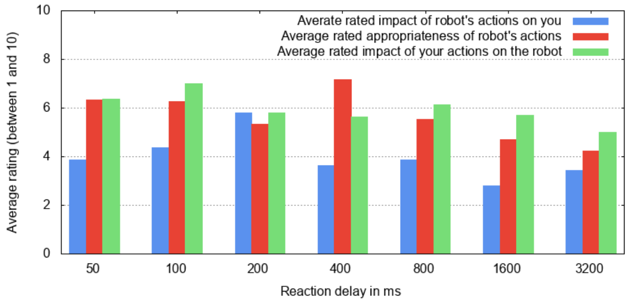

Fig. 4. The resulting average ratings in relation to the delay time in milliseconds concerning the rated impact of the robot's actions, the impacts participants made on the robot, and the rated appropriateness of the robot's actions.

highest. The resulting averages for the last two questions can be seen in Figure 4.

Grouping the results by the age of the participants shows that most of the age groups prefer human-like reaction times. The top-rated of the average reaction time for the affective impact of the robot's behavior in regards to age group was 200ms. Our initial assumption was that the results would support a relationship between older age and slower preferred reaction times. However, this is not the case. The 200ms delay which corresponds to human-like reaction times is preferred even by the older test participants. The age group from 21 - 30 preferred the slowest reaction time of 3200ms, but a closer look at the data reveals that may be explained by a lack of proper age distribution for some delay times. It is vital to state that the age distribution across every delay group is not uniformly distributed. Some delay categories have very few examples for specific age groups. The results indicate that the age group of 41 - 50 prefers a slower-than-human robot with a preferred reaction time of 400ms and presents an opportunity for further research projects to focus more on each age group and the preferred reaction times.

VI. FROM MEASURES TO MEANING

The boundaries of each discrete state in models such as the PAD space are fuzzy, and a single 3d coordinate can rarely convey the rich sources of information that affective data is [38]. Because emotions are given significance by the words that express them, they differ between languages. In some cases with specific languages, certain emotions are not present or mean something different [46]. When the interpretation and comprehension of the affective states are culturally dependent the problem is that the interpretation of them change with each cultural context and group of human observers [47]. This paper acknowledges that it is difficult to create a test setup that provides clear answers, but attempts to work around it by using many participants. Our test setup had the two following drawbacks regarding the age of the participants:

1. The test was designed to measure the effect of the delay times. This meant that the age groups were not uniformly distributed within each tested delay times

and that some delay times had one or more age groups that were not represented.

2. As our delay time was doubled each time, it left out too many details of the interesting area between 200 and 400ms. It may be that the effect we were attempting to verify was smaller than anticipated and that we instead needed a test that expanded the knowledge on that specific delay interval.

The results of the first test indicate that there is a causality between the level of perceived arousal and the coordination of human-robot reactions to the context event. The robots were perceived as being more aroused when their reactions were coordinated with human observers. The results show that considering the timing aspects of conveying affective information and sharing a reaction with a human observer can be beneficial in those scenarios where the aim is to convey highly aroused affective states.

That the overall voted most suitable reaction delay time for the reaction to the physical interaction of the second test is 200ms, might for some scenarios be considered a positive result. Such a delay leaves a wide timeframe even for low hardware-driven robots to analyze the input and consider the proper reaction to a given situation. The physical properties in the second test also seemed to affect how the participants interpreted the overall pleasantness of the interaction. Some participants stated they felt bad about hitting the robot and did not want to interact with it because it seemed as if they punished the robot for no reason. The average ratings on appropriateness in relation to reaction time can be seen in Figure 4.

The resulting ratings for the different delay times fortify what Pan et al. 2019 found with humans and robots interacting in a cooperative setting [37]. Our common intuition would say that the Pan et al. test participants preferred a human-like response time because they used a humanoid robot and a human-to-human inspired context with a cooperative task. However, if we interpret the highest-rated suitable behavior as the preferred behavior, our result shows that these findings can be extended to non-humanoid robots as well. They also show that the same reaction time was found most suitable in high-intensity scenarios - in which people physically interact with the robot.

When we asked the participants to rate the emotional impact of hitting the robot, the highest average rating was given when the robot reacted with a delay time of 100ms followed by the second-highest ratings for 50ms. This could indicate that there is a measurable difference between how the participants wanted the robot to react in the different scenarios. When the aim is to convey to the participants that their actions had a large impact, the reaction time should be shorter than human reaction times (<250ms). When the aim is for the robot to make a large emotional impact on the participants, the robot should react similarly to humans (~250ms). It makes sense to consider to what extent the results are applicable in other contexts. The tested scenario portrayed a social context, and it may be that the highest-rated reaction speeds in this experiment would be found suitable for other social situations as well. However, the results do not per se extend to other robot types and or other domains. E.g. we don't necessarily prefer a manufacturing robot at a factory to work at the same speeds as humans.

Regarding the results grouped by age, we argue that the presented findings introduce many opportunities for further research on the topic. As one, we suggest investigating more specifically to what extend age influences the chosen most suitable reaction times in a finer interval between 200 and 400 ms - to see if age specific reaction times could strengthen the reception of affective information even further.

VII. Conclusion

The paper has investigated the causality between coordinating human-robot reactions and the perceived affective impact on robots. It has shown that we can use coordinated reactions to strengthen the way robots convey affective information. The emphasis was to see whether the perceived level of intensity in the behavior was increased when a robot was reacting to a context event in coordination with a human, and to test whether delaying the specific reaction times in physical interactions influenced how test participants viewed the affective state of the robot. We carried out two human-robot interaction tests to highlight these aspects of human-robot interaction.

The result showed that there was a significant difference between how aroused the human observers rated the robots as being in the first test when the human-robot reactions were coordinated. The results of the second test indicated that even for high-intensity scenarios with non-humanoid robots, the preferred reaction for the robots was similar to the reaction time of humans. Furthermore, they showed that a faster reaction time (~100ms faster) was preferred when the goal was for the test participants to feel as if they made a large impact on the robot.

The findings indicate that the concept of sharing reactions and using near-human reaction delays can be strategically used to influence how the current affective state of a robot is perceived.

References

[1] K. S. Morten Roed Frederiksen, "A systematic comparison of affective robot expression modalities," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019.

[2] G. Hoffman, M. Cakmak, and C. Chao, "Timing in human robot interaction," *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 509–510, 2014.

[3] L. Boccanfuso, E. S. Kim, J. C. Snider, Q. Wang, C. A. Wall, L. DiNicola, G. Greco, F. Shic, B. Scassellati, L. Flink, S. Lansiquot, K. Chawarska, and P. Ventola, "Autonomously detecting interaction with an affective robot to explore connection to developmental ability," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7, 2015.

[4] W. D. Stiehl, C. Breazeal, K.-H. Han, J. Lieberman, L. Lalla, A. Z. Maymin, J. Salinas, D. Fuentes, R. L. Toscano, C. H. Tong, A. Kishore, M. Berlin, and J. Gray, "The huggable: a therapeutic robotic companion for relational, affective touch," *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.*, vol. 2, pp. 1290–1291, 2006.

[5] A. Singh and J. E. Young, "A dog tail for communicating robotic states," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 417–417, 2013.

[6] C. C. Bennett and S. Sabanovic, "Perceptions of affective expression in a minimalist robotic face," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 81–82, 2013.

[7] D. Benson, M. M. Khan, T. Tan, and T. Hargreaves, "Modeling and verification of facial expression display mechanism for developing a sociable robot face," *2016 International Conference on Advanced Robotics and Mechatronics (ICARM)*, pp. 76–81, 2016.

[8] C. L. Lisetti, S. M. Brown, K. Alvarez, and A. H. Marpaung, "A social informatics approach to human-robot interaction with a service social robot," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, 2004.

[9] G.-Y. Park, S.-I. Lee, W.-Y. Kwon, and J.-B. Kim, "Neurocognitive affective system for an emotive robot," *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2595–2600, 2006.

[10] C. Breazeal, "Emotion and sociable humanoid robots," *Int. J. Hum.-Comput. Stud.*, vol. 59, pp. 119–155, 2003.

[11] R. Gockley, R. G. Simmons, and J. Forlizzi, "Modeling affect in socially interactive robots," *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 558–563, 2006.

[12] S.-I. Lee, G.-Y. Park, and J.-B. Kim, "Natural emotion expression of a robot based on reinforcer intensity and contingency," *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 2144–2149, 2007.

[13] J.-D. Yim and C. Shaw, "Designing cally, : a cell-phone robot," in *CHI Extended Abstracts*, 2009.

[14] J.-Y. Yang, B. K. Dinh, H.-G. Kim, and D.-S. Kwon, "Development of emotional attachment on a cleaning robot for the long-term interactive affective companion," *2013 IEEE RO-MAN*, 2013.

[15] G. Yoshioka, T. Sakamoto, and Y. Takeuchi, "Inferring affective states from observation of a robot's simple movements," *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 185–190, 2015.

[16] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 53–60, 2010.

[17] C. L. Bethel, K. Salomon, and R. R. Murphy, "Preliminary results: Humans find emotive non-anthropomorphic robots more calming," *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 291–292, 2009.

[18] C.-Y. Lin, L.-C. Cheng, C.-C. Huang, L.-W. Chuang, W.-C. Teng, C.-H. Kuo, H.-Y. Gu, K.-L. Chung, and C.-S. Fah, "Versatile humanoid robots for theatrical performances," *International Journal of Advanced Robotic Systems*, vol. 10, p. 1, 01 2013.

[19] B. Gonsior, S. Sosnowski, M. Buß, D. Wollherr, and K. Kühnlenz, "An emotional adaption approach to increase helpfulness towards a robot," *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2429–2436, 2012.

[20] I. D. Addo and S. I. Ahamed, "Applying affective feedback to reinforcement learning in zoei, a comic humanoid robot," *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pp. 423–428, 2014.

[21] K. Winkle and P. Bremner, "Investigating the real world impact of emotion portrayal through robot voice and motion," *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 627–634, 2017.

[22] C. Becker-Asano and H. Ishiguro, "Laughter in social roboticsno laughing matter," *Intl. Workshop on Social Intelligence Design (SID2009)*, 01 2009.

[23] J. H. Lui, H. Samani, and K. yu Tien, "An affective mood booster robot based on emotional processing unit," *2017 International Automatic Control Conference (CACS)*, pp. 1–6, 2017.

[24] R. Zhang, Y. He, and Z. Liu, "Communication atmosphere in humans and robots interaction based on fuzzy analytical hierarchy process," *2017 36th Chinese Control Conference (CCC)*, pp. 6767–6772, 2017.

[25] I. Cohen, R. Looije, and M. A. Neerincx, "Child's recognition of emotions in robot's face and body," *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 123–124, 2011.

[26] T. L. Chen, C.-H. King, A. L. Thomaz, and C. C. Kemp, "Touched by a robot: An investigation of subjective responses to robot-initiated touch," *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 457–464, 2011.

[27] J. Xu, J. Broekens, K. V. Hindriks, and M. A. Neerincx, "Effects of a robotic storyteller's moody gestures on storytelling perception," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 449–455, 2015.

[28] J. Xu, J. Broekens, K. Hindriks, and M. A. Neerincx, "The relative importance and interrelations between behavior parameters for robots' mood expression," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 558–563.

[29] W. Johal, S. Pesty, and G. Calvary, "Toward companion robots behaving with style," *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, vol. 2014, pp. 1063–1068, 08 2014.

[30] L. Rincon, E. Coronado, H. Hendra, J. Phan, Z. Zainalkefli, and G. Venture, "Expressive states with a robot arm using adaptive fuzzy and robust predictive controllers," *2018 3rd International Conference on Control and Robotics Engineering (ICCRE)*, pp. 11–15, 2018.

[31] C. Park, J. Ryu, J. Sohn, and H. Cho, "An emotion expression system for the emotional robot," *2007 IEEE International Symposium on Consumer Electronics*, pp. 1–6, 2007.

[32] C. L. Bethel and R. R. Murphy, "Survey of non-facial/non-verbal affective expressions for appearance-constrained robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, pp. 83–92, 2008.

[33] J. Michael, "Shared emotions and joint action," *Review of Philosophy and Psychology*, vol. 2, pp. 355–373, 01 2010.

[34] V. V. Unhelkar, P. A. Lasota, Q. Tyroller, R.-D. Buhai, L. Marceau, B. Deml, and J. A. Shah, "Human-aware robotic assistant for collaborative assembly: Integrating human motion prediction with planning in time," *IEEE Robotics and Automation Letters*, vol. 3, pp. 2394–2401, 2018.

[35] L. Lucignano, F. Cutugno, S. Rossi, and A. Finzi, "A dialogue system for multimodal human-robot interaction," in *ICMI*, 2013.

[36] R. A. Hyman, "Stimulus information as a determinant of reaction time." *Journal of experimental psychology*, vol. 45 3, pp. 188–96, 1953.

[37] M. B. Matthew Pan, Espen Knop and G. Niemeyer, "Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019.

[38] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," in *Face and Gesture 2011*, March 2011, pp. 827–834.

[39] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer, "Human-robot interaction in handing-over tasks," *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 107–112, 2008.

[40] R. W. Bing and J. Michael, "Overcoming the uncanny valley through shared stressful experience with a humanoid robot," *Journal of Emerging Investigators*, 06 2018.

[41] M. Mori and K. F. MacDorman, "The uncanny valley: The original essay by masahiro mori," in *IEEE Spectrum*, 2017.

[42] A. M. R. von der Pütten, N. C. Krämer, and J. Herrmann, "The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior," *International Journal of Social Robotics*, vol. 10, pp. 569–582, 2018.

[43] A. Mehrabian and J. Russell, "An approach to environment psychology," *Cambridge, MA*, 01 1974.

[44] J. A. Russell, "Core affect and the psychological construction of emotion." *Psychological review*, vol. 110 1, pp. 145–72, 2003.

[45] L. V. Moshkina, S. Park, R. C. Arkin, J. K. Lee, and H. Jung, "Tame: Time-varying affective response for humanoid robots," *I. J. Social Robotics*, vol. 3, pp. 207–221, 2011.

[46] B. B. de Mesquita and N. Frijda, "Cultural variations in emotions: a review." *Psychological bulletin*, vol. 112 2, pp. 179–204, 1992.

[47] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," *IJSE*, vol. 1, pp. 68–99, 2010.

13

Paper 3: Adaptable context-based behavior selection in autonomous robots

Adaptable context-based behavior selection in autonomous robots

Morten Roed Frederiksen[1] Kasper Stoy[2]

Abstract— **We all adapt to our current physical context when we communicate with other people. We lower our voice or movement speed in small spaces or increase them in larger physical environments. For robots to partake in social interaction with humans, adapting their behavior to the context should be a requirement. This paper presents a novel system that uses contextual knowledge to guide a robot's behavior in human-robot interactions. The system consists of two parts: one that represent previously encountered contexts and one that through human-robot interaction learns to prioritize behaviors in each of them. The contexts are identified and clustered by their physical proporties. In a context the system autonomously tests different behaviors and learns to prioritize between them to match the users preference and the contextual information. The system was tested using a custom built affective robot in 2 different physical contexts, with 4 distinctive behaviors through 72 interactions with 6 different test participants. The system managed to adjust its behaviors to match the physical context in accordance with the participants. A high intensity behavior was generally found fitting for the largest of the two context. The result is statistically significant at ($p<.05$). Social robots should be able to adapt to the context and we have shown that while a richer context distinction may be preferred for context-guided robot behavior selection, with simple means, a single attribute can be viable to drive the selection with a significant confidense.**

I. Introduction

The physical context of our interactions have a significant influence on our behavior. We adapt to it and moderate different aspects of our behavior with each context we encounter. E.g. a large physical environment may allow us to be physically more engaged in interactions while a smaller space demands that we are less animated. For social robots, the context poses a challenge for them to communicate effectively. The physical environment, the proxemics, the audio volume, and the amplitude of its movements are all examples of adaptable parameters that can enhance an interaction. If these are not attuned to the physical demands of the context the robot may not be able to communicate effectively. Eg. a loud robot can be heard at all times but would probably be a terrible fit for more intimate interactions.

When social robots are designed they are often equipped with a few behaviors to handle a variety of social interactions. These behaviors often work well in the context they are designed to work in and encourages a positive anthropomorphic interpretation [1], [2]. E.g. among other behaviors, the Vector robot by Anki has behaviors to map its surroundings and entertainment behaviors that make it dance or tell jokes [3]. While such robots work very well in specific contexts, placing them in a different context can make some of their actions seem inappropriate (while other behaviors may fit the changed context). A preconfigured robot without adaption will unknowingly use the same behaviors even though they have no or even negative effect on the interaction in a changed context. Both the social, cultural, and physical contexts are important for the robots because we interpret their actions in light of the context in which they are performed [4].

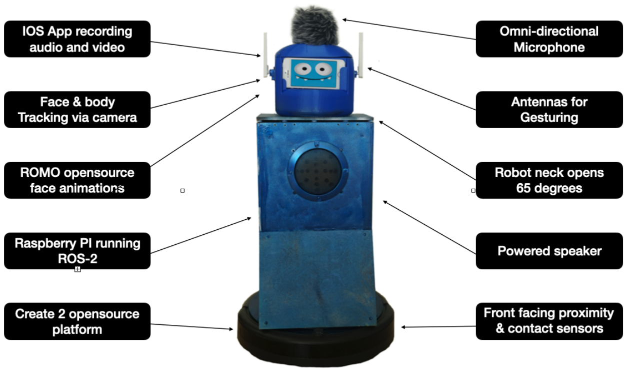

Fig. 1. Affecta V3, The humanoid robot we created to test the system.

Adapting behaviors to the context can be a drive for more diverse behavior choices in such projects, and developing context aware systems may be the hurdle to overcome to unlock better social functionality of robots outside of lab conditions [5], [6].

A. Previous approaches

There have been a few projects on optimizing robot capabilities according to the physical context. Narayanan et al. 2011 created movements based on visual perception of the environment and Jamone, Damas, and Santos-Victor 2014 created dynamic mapping models based on interactions with various objects [9], [10]. The models in the latter allowed the robot to approximate the torque rate for correctly interacting with different context objects. Pandey et al. 2010 created a framework that paid special attention to humans as objects in the vicinity as their navigation system analysed local clearance and environment structure [11]. Navigation in social spaces was also the focus of Banisetty et. al 2019 in a system that among other factors considered social conventions and physical obstacles to guide promixity constraints for the robot's navigation path around humans [6]. Torre et al. 2020 investigated the alignment of voice for the specific context,

[1]Morten Roed Frederiksen mrof@itu.dk and [2]Kasper Stoy ksty@itu.dk are affiliated with the REAL lab at the Computer science department of The IT-University of Copenhagen, Rued Langgaards vej 7, 2300 Copenhagen S

task and robot morphology. They tested different voices and let users select the voice that matched a robot and a physical context. They found that the task context was a main driver for which robot was chosen for a voice [12].

The acoustic properties of environments was used in Lera et al 2017 to classify indoor contexts [13]. The project used convolutional neuronal networks to classify different contexts based on ambient audio signals. Cosgun & Christensen 2018 added context awarenes to enhance a person following robot. The robot predicted targets of future human interaction by looking at the velocity of the human it was following [14]. Liu, Wang, and Wang 2018 used context aware pose estimation to ensure safety for users in the vicinity of their assembly robot. The pose estimation system used a kinect, leap motion and MYI sensor to recognize the intentions of the user by classifying key assembly poses [15]. Xiao et al. attempted to increase the contextual knowlede of a robot in an interaction by allowing communication through natural body language. The robot they created could understand the meaning of human upper body gestures and would communicate by using movements, facial expressions, and verbal language [16]. Robots that autonomously determines personality traits of the users were the focus of Zafar et al. 2018 and Zafar et al. 2019. Their solution used speech charactiristics and found promising resuts in detecting personality traits [17], [18]. Their approach used excerts from the interactions annotated by a psychologist to link non verbal cues to level of exhibited extroversion.

A few of these are examples of how robots dynamically alter control in conjunction with immediate or processed sensor input. But for robots that interact with humans in unknown physical context, the task may be more complex. If we know the physical context ahead of an interaction it is viable to pre-define the best contextual behaviors. However, this is often not the case. There is a potential in learning because solely relying on pre-selected attributes to distinguish contexts may not provide enough information. Eg. there is no guarantee that the thresholds for physical proportions for a specific behavior are valid for every scenario. The potential in dynamically adapting behaviors to the physical context is that it may allow placing robots in previously unvisited physical contexts and it would then learn to optimize its behavior to fit them.

The current affective status of a human describes how the human experiences its current set of emotions and sensory information (a simplified explanation is that it is how we feel about what we current see, hear and think about). The expected affective impact (how much the current interaction changes the current affective status) of different behaviors is difficult to know ahead of an interaction, making it difficult to preconfigure the behavior-selection of a robot for optimal affective impact in a specific context. The actual impact may vary with different robot morphologies, different cultural contexts, etc. which is why there may be unused potential in learning the affective impact along with the discovery of new contexts.

B. Context-based behavior selection

In this paper, we create and test a minimalistic solution for context-specific behavior selection, a novel robot system that learns to prioritize between a set of preconfigured discrete behaviors for a specific physical context. This is a combination of a preconfigured behavior approach and a dynamic learning approach. The robot we have created for this purpose was not configured to work in a specific context but instead optimizes its behavior over multiple previously unexplored contexts. The robot gradually explores and learns the specifics of new contexts and attempts to place them in a topography of nodes representing existing previously know contexts. As more contexts are visited the robot generalizes on the input and clusters similar context nodes in the vicinity of each other. Each node represents designated behavior strategies that fit this type of context. As the robot encounters humans it interacts with them and attempts to verify the impact of its different affective behaviors and to update the priority in the behavior set that matches the current context.

The robot was tested in 72 different interactions with 6 human participants using a refined version of our custom-created affective robot "Affecta" as seen in Figure 1. The details of the robot are explained in the following sections. The tests we performed showed that our robot was able to distinguish between individual contexts and that it created a prioritized set of behaviors that adapted to the users' preference in accordance with the physical context. As expected, a high intensity level behavior was found more fitting in a physically large room of 6x5m, while a lesser intensive action was found more appropriate in a smaller room of 2x3m with a significant majority at $p<.05$.

Although the robot system we present in this paper is a simplification of complex human affective processes, the results indicate that it can provide a good foundation to build upon for projects on non-preconfigured behavior control. Even with limited sensor input in our test scenario, the robot yielded a clear context-aware prioritization. We opted to focus on the prioritizations of four discrete behaviors, but in further research projects, the results indicate that there is no reason not to apply learning on a much larger set of context-specific behaviors. Based on our initial findings, we conclude that the system allows for dynamic (simplified) behavior selection and we suggest focusing on implementing similar systems for personality-based behavior selection for future improvements.

II. METHOD

We used the robot depicted in Figure 1 to autonomously distinguish between different physical contexts and to find the best possible behavior for each of them. We did this by letting the robot collect information about its current context and by letting the robot interact with humans asking their opinions on the different behaviors.

A. Representing a context

The robot we created for this project was a 75cm tall humanoid robot using the open source "Romo" face project

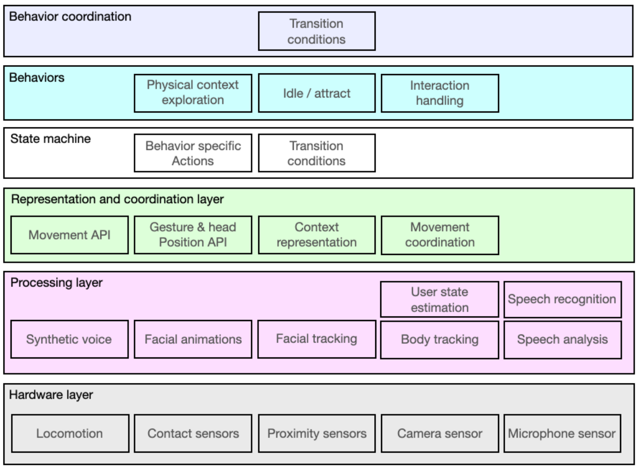

Fig. 2. Diagram of the layered system architecture of the Affecta V3 robot we created for the tests.

and built on the “iRobot Create2” platform [19], [20]. The Create2 platform consisted of an open source version of the Roomba robot. It provided a foundation to build upon plus sensors, and actuation. The robot had three contact sensors placed at the front bumper on the robot. It also had three low range distance sensors, one at the front and one on either side. For actuation the robot used two electric motors in a differential drive setup allowing the robot to move forward, backward, and to turn around its own center axis. The architecture diagram of the Affecta robot can be seen in Figure 2.

The robot used an estimated physical size of the environment as the sole attribute defining the physical context. Determining the room size was achieved by measuring time-of-drive in the physical context, using the front facing contact- and proximity sensors on the robot to determine time of contact. The robot would move around in each physical context and measure the time between each collision with either walls or objects in the environment. As the robot detected a collision it would turn around randomly between 140-270 degrees and start another measurement. Five samples would be averaged and a vector containing this single value would represent the current context. Estimating the dimensions of a physical context can be done easier with a richer sensor setup. We used time-of-drive as the method as it is available to most mobile robots using simple contact sensors.

Technically each contex can be represented by multiple averaged environment samples in that same vector. With richer sensors available other physical characteristics could be sampled to give a more precise representation of the current context in each vector. The calculated distance between individual context-vectors dermines the similarity of their physical properties. If the sum of squared difference between a preexisting context vector and a newly sampled context vector is below a predefined threshold, the current context is assumed to be the same context that the pre-existing vector represents.

B. Updating the representation

A vital part of the system is a representation of the different contexts in a data structure that provides dimensionality reduction for the gathered context vectors. The data structure we used is a self organizing map that is able to cluster similar contexts in the topological vicinity of each other [21], [22]. Each of these topological positions match small variations in different contexts. We altered our self organizing map implementation to provide emphasis of individual node weights (see details below) and to provide a direct input-to-topological-position method.

In our context represenation we have created a 10x10 matrix of contexts context-vectors with a 2d position for each entry. The matrix is initialized with a context vector for each position and the values of these vectors are randomized between 0 and 1. Each time new contexts are explored and new context-vectors are created, the most similar context-vector in the matrix is retrieved and each measure point in this context-vector is updated by the difference between each value modified by a fixed learning rate. The distance between context-vectors is defined as the sum of squared difference between all value paris of the vectors. As the closest matching contet-vector is updated, the nearest contexts in euclidean distance around it are also updated with a learning rate that decreases (halves) with each distance step away from the center context-vector. In similar style as self organising maps, introducing a fundamentally different new input vector automatically creates a new region in the topology of the 2d map by altering the existing context-vectors.

In our implementation, the distance between context-vectore is furthermore calculated with attention to the importance of each of the gathered mesurements in it. Some measure points may be more important than others and could have a greater impact on similarity when determining the distance between contexts. To model this, each measure point in a context-vector has an added importance modifier that multiplies the difference value. In this paper, the context is identified by a single important measure points so the modifier for that attribute is set to 1.0.

C. Behaviors in a context

In this paper we investigate the robot’s ability to prioritize between four different predefined behaviors to dynamically match the current context. As such, all contexts in our 2d matrix have a set of behaviors attached. These are ordered in priority after the best fit for the context. The behaviors consist of series of movement (forward, backward, and tuning movements), head antenna gesturing (waving pattern), facial expression animations, and audio expressions. Each behavior differs in audio and animations but shares the same movement and gesturing patterns at different intensity levels from 0 to 3, meaning that each of the four behaviors has a unique intensity level. The intensity level defines the size of the physical movements and gesturing, and also determines

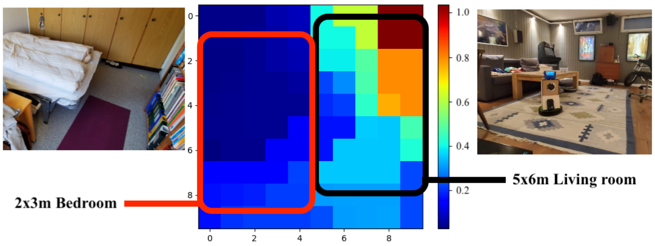

Fig. 3. The physical context representation after 5 minutes of exploration in each context. The 10x10 square map represents 100 individual contexts, and the heat map colors represent the normalized time-of-drive measurement values. The highlighted areas show how the measurements gathered in each context has created a nieche area for the two contexts in the self-organizing-map representation.

the intensity of the animated expression. The highest intensity level at three has the largest movement and gesturing actions while intensity zero does not have any movements or gesturing but instead consists solely of animations and audio expressions. The robot prioritizes between behaviors of the context by testing them through interactions with humans, and by gradually adjusting its prioritization for the indentified physical context of each encounter. The behavior priority for each context is updated with the same strategy as the physical context-vectors with the behaviors of the neighburing contexts being updated with a decreasing learning rate as well. The topology of the nodes may be altered when new contexts are discovered which indirectly influence the behaviours prioritized in similar contexts. Further descriptions of the interactions are decribed in the following section.

III. The experiments

The robot was tested in two different experiments. The first experiment aimed to verify that the system was able to autonomously distinguish between the contexts using a single attribute and create a discernable representations for each of the indentified contexts. The second experiment tested the affective impact of the four different levels of intensity behaviors through interactions with human participants, and verified that the robot created behavior priorities that coupled the physical contexts with matching behavior intensities.

A. Experiment One: Physical context exploration

The robot was designed to work autonomously through all phases. The first of which included measuring the context. We tested the robot in two different contexts, a 6x5m living room, and a 2x3m bedroom context. It gathered measurements for five minutes in each room and simultaneously updated its context representation with each gathered data point. The created context representation was visualized in relation to the average time of drive measure point and can be seen in Figure 3.

Before initiating any interactions with test participants the robot would recognize its current context by moving around in the physical context and gathering three measurements. This was repeated before every new interaction.

The robot performed exploration to identify the current physical context. This was achieved by moving around until the robot had performed 3 successful measurements of time-of-drive in the physical space.

1. The measure was recorded as the time between either of the front-facing distance sensors detected an object in close proximity or that the front-facing contact sensors detected a physical collision. Any detection within 1 second apart from the last previous measurement would be ignored to hinder unprecise subsequent measurements.
2. Once the robot detected anything in front of it, it would turn in a random direction between 120 and 240 degrees. If the robot got stuck on any obstacles we would step in and untangle it.
3. The value was averaged between the 3 successful measurements and the context of the closest matching distance was retrieved and the robot would stay idle.
4. The robot would stay idle in the same position until it detected a human to interact with (detected using the body and facial recognition). while it idled, the robot would rotate slightly once every 15th second to gain a larger field of view in the context.

B. Experiment Two: Behavior prioritization

To verify the robot's ability to adapt behaviors to the context via human interactions, we tested it through interactions with 6 different test participants aging from 9 to 70. The tests were performed in two different physical contexts. The main attribute we considered in this system was the physical room size and therefore we tested the system in two different sized locations: a small indoor bedroom room (2x3m) and a larger living room (6x5m). Although the two spaces had very different dimensions they shared acoustic properties and noise levels. The test participants interacted individually with the robot one after another. Each interaction lasted for about ten minutes. The participants were informed that the robot would interact with them as soon as it could see them. They were not instructed to perform any specific actions other than to interact with the robot and answer truthfully to any questions it would ask them. The following steps were performed with each participant:

1. When the robot detected the visible face of a participant in the vicinity, the robot would initiate the interaction by asking the participant: *"Hello there, can I ask you a question?"*.
2. If the participant approved, the robot would ask *"Could I ask you to tell me a bit about yourself, like who are you and what do you do?"*.
3. The robot would listen while the participants answered and continue by saying *"Okay.. thank you, that sounds nice. Listen I actually need your help.. I would like you to assess my behavior.. That's really really difficult for me and I would appreciate your honesty. I will show you two different behaviors and I would like you to compare the two. Are you ready? Here is the first.."*.

4. The robot would select two random behavior from the current context and show the first of them to the participant.
5. The robot would say *“Okay.. watch carefully.. here comes the second behavior”*. After which the robot would show the second selected behavior.
6. The robot would continue by asking *“so.. please be honest, which of the two behaviors did you like the most?”*
7. The robot would then ask *“And which of the two behavior did you find most fitting in this context?”*

Fig. 4. A test participant interacting with the robot in the larger 6x5m physical context.

The robot handled all interactions with the test participants including obtaining an answer on preferred behaviors. Using a robot to ask the questions retains some of the benefits of a face-to-face interview rather than using a questionnaire after the interaction which may introduce bias on self-report questions as found in Heath et al 2020 and Althubaiti et al. 2016 [23], [24]. The test interaction can be seen in Figure 4.

The behavior selection strategy for testing the different behaviors in the interactions was similar to a classic epsilon greedy reinforcement approach with one part exploration and one part verification (in a 9/1 ratio). The exploration path used a random behavior while the verification strategy would use and test the currently rated best-fitted behavior. In each interaction two behaviors were compared and the participants were asked a question regarding the test behavior’s suitability in the current physical context. This resulted in the two behaviors being found either suitable or not-suitable for the physical context. Each behavior was updated following the interaction and the sum of positive votes out of the total number of votes would define the behavior-rating for that specific physical context going forward. The behaviors that were used as comparison in the interactions were randomly selected with each encounter. During the full range of tests, 72 behaviors were tested across the two contexts and 96 physical context measurements were gathered.

IV. Results

Figure 5 shows the percentage of total positive votes from all interactions for each behavior intensity in the different physical contexts. There is a significant ($p<.05$) difference in the distribution of votes for the two physical contexts when the robot asked the users to chose the most fitting of the two displayed behaviors. The found most fitting behavior for the largest (6x5m) context was the behavior with intensity level 2 (selected in 73% of times tested), while the lowest rated behavior from the same context was the behavior with intensity level 0 which was only found most fitting in 20% of the times it was tested. For the smaller physical context (2x3m) the behavior found most fitting was the behavior with intensity 1 with 66.7% positive votes, while the least fitting behavior was the behavior with intensity level 3 27.3% of the positive votes.

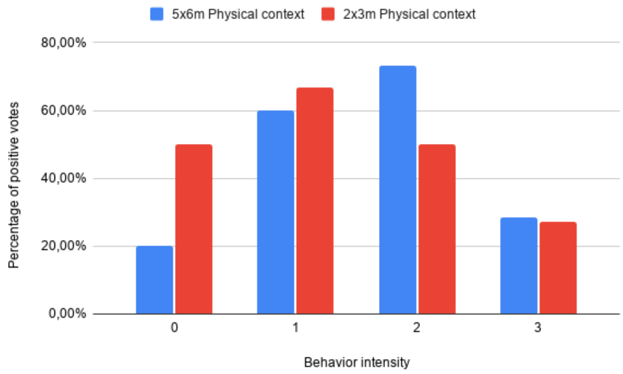

Fig. 5. The percentage of positive ratings for each behavior in the two physical contexts. The blue series depicts the largest (6x5m) physical context, while the red series depics the smaller (2x3m) physical context. The x-asis shows the different behavior intensity levels while the y-axis shows the percentage of positive votes for each behavior.

The robot updated its behavior prioritization in accordance with the interaction results and the visualization can be see in Figure 6. The visualization depicts the two main physical context regions and the voted most fitting behavior intensity in each of them with an intensity level of 2 (yellow) for the 6x5m physical context and an intensity level of 1 (light blue) for the 2x3m physical context. The maximum intensity level of 3 is found mainly in the upper right region of the context representation which matches the largest found physical context measurements in Figure 3. Comparing the physical context visualization and the context behavior representation reveals that the adjusted behavior intensity levels intuitively matches the physical aspects of each context but also reveals that the measurements performed prior to each interaction often places the robot between two well defined context regions in the representation.

Summing up the results there are three main findings. The first finding is that the robot managed to distinguish between each context using a single physical context attribute to describe the context. The second is that the robot managed to find a statistically significant prioritation for behavior intensity for each context. The third finding is that the robot adjusted the behavior prioritation for each context in accordance with the physical attributes of each context.

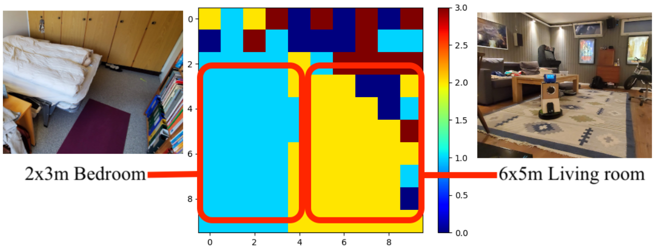

Fig. 6. The context behavior representation. The 10x10 map represents the 100 individual contexts, while the heat map color represents the highest prioritized behavior intensity of each context. The light blue color is intensity level 1 while yellow color is intensity level 2. The topological regions are similar to those in Figure 3.

V. Applying the Findings

This paper aimed to investigate the feasibility of using physical context information as a guide to drive behavior selection for a robot. There is a definite and clear relationship between the physical properties of the room and the human-preferred physical properties of the robot's behaviors. The results indicate that as the physical space increases around the robot, and the participants who interact with it, so does the amount of preferred intensity in the robot's behaviors. The robot managed to create regions in its context representation for each physical context and update the same regions with a behavior that matched the increased physical dimensions. However, the system did not map the context consistently to the same node in the context representation. This stems from the uncertainty of using a single attribute to recognize the current context. Adding further parameters to identify the physical aspects of the contexts would make it more precise. However, there is a strength and a point in using a single attribute, as it is easy to apply to most robots and it provides contextual information even from simples sensors.

The system adjusts its behaviors using a single context attribute This is a simplification of how behaviors depend on context in humans which is a complex psychological process. As it is making assumptions on a sparse set of information it will often be wrong when determining contexts and best behaviors. However, the same can be said for most humans. We don't always find the completely correct behavior for a given situation, we adjust our immediate behaviors and negotiate affective status many times through each interaction [25]. Many different sensor types could have been used to provide contextual information about the physical context, but we aimed at using simple sensors that would often be available in off-the-shelves robots to make the system applicable in a large variety of other robot projects. For that reason, we opted to focus on room size as the simple context attribute that influenced how a robot should behave. We also experimented with using reverb decay to indicate the room size as a second attribute, but we choose to disregard that attribute to emphasize the simplicity in driving behavior selection using only time-of-travel as the input. More complex and richer sensors could provide better distinctions between contexts but may also introduce unusable sensor data. Some attributes aid behavior selection better than others. E.g the detected color of various rooms would make it easy to recognize a known context, but that attribute says nothing about the optimal behaviors. The kind of behaviors that the system optimizes for determines the choice of attributes, meaning that it is a tradeoff between context distinction and more generalizable behaviors.

The last two behaviors with intensity 0 and 3 were mostly down prioritized. Outside of them not generally fitting the context, this could also be due to the nature of the interaction. The participants interacted with the robot from a distance between 1-2 meters and a large physical behavior (intensity 3) might seem intimidating comming from a robot, and a lack of movement (intensity 0) might make the robot seem inanimate or uninteresting. However, some people preferred these behaviors and it indicates that personal preference also plays a part in each interaction. The individual mood and personality of each test participant may also influence how each behavior is received.

When we asked people for their reasoning behind their opinion on the behaviors tested in the interaction, a lot of them gave very anthropomorphic reasons for their choices projecting human reasoning and motives behind the robot's behavior. Eg. "the robot rudely changed the flow of the conversation with some moves", or "the robot did not really listen to what I told it, so I got a bit annoyed with it". This further indicates that the individual personality of the users had an impact on how they viewed the behaviors. Some previous research projects have already indicated that personality traits such as extroversion influence our preference for robot behaviors. Future research directions from this project could benefit from investigating context based behavior selection in combination with more advanced awareness of user traits.

VI. Conclusion

This paper investigated the abilities of a system that enables a robot to identify physical contexts, and autonomously test and prioritize between predefined behaviors in each of them. The system was tested in two different physical contexts through 72 interactions with 6 human participants. The ability to identify contexts was verified as working successfully using a single measurable parameter to distingush the physical properties of each environment. The system created distinguishable regions for each physical context in its context representation. The robot also managed to autonomously prioritize behaviors in each identified context. The resulting list of behaviors matching each physical context were learned through interactions with the users, and the viualized map of behaviors showed that the corresponding behaviors had been correctly prioritized by the robot as it adapted to each context. The results highlight the potential in minimalistic context-based behavior selection and demonstrate the possibility of basing context-based behavior selection on the information retrieved from a single measurable attribute such as averaged time-of-drive. The people who interacted with the robot had a tendency to project anthropomorphic reasoning behind the robot's behavior and we recommend proceeding with an extension to this system which allows for greater behavior flexibility based on the atttributes of the users rather than solely the context.

REFERENCES

[1] W.-R. Ko and J.-H. Kim, *Behavior Selection Method for Entertainment Robots Using Intelligence Operating Architecture*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 75–84. [Online]. Available: https://doi.org/10.1007/978-3-642-37374-9_8

[2] M. Waytz, A, E. C.K, M. N, G. G, and J. T. Cacioppo, "Making sense by making sentient: Effectance motivation increases anthropomorphism." in *Journal of Personality and Social Psychology*, 2010, pp. 410435, 99(3).

[3] Anki, "Vector: The home robot with interactive ai technology," 06-20-2020. [Online]. Available: www.anki.com/en-us/vector

[4] R. Read and T. Belpaeme, "Situational context directs how people affectively interpret robotic non-linguistic utterances," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 4148. [Online]. Available: https://doi.org/10.1145/2559636.2559680

[5] M. Luber, L. Spinello, J. Silva, and K. Arras, "Socially-aware robot navigation: A learning approach," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 10 2012, pp. 902–907.

[6] S. Banisetty, S. Forer, L. Yliniemi, M. N. Nicolescu, and D. Feil-Seifer, "Socially-aware navigation: A non-linear multi-objective optimization approach," *ArXiv*, vol. abs/1911.04037, 2019.

[7] K. S. Morten Roed Frederiksen, "Robots can defuse high-intensity conflict situations," in *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robot and Systems (IROS)*, 2020.

[8] M. R. Frederiksen and K. Stoy, "Augmenting the audio-based expression modality of a non-affective robot," *International Conference on Affective Computing and Intelligent Interaction IEEE*, 2019.

[9] K. K. Narayanan, L. F. Posada, F. Hoffmann, and T. Bertram, "Scenario and context specific visual robot behavior learning," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 1180–1185.

[10] L. Jamone, B. Damas, and J. Santos-Victor, "Incremental learning of context-dependent dynamic internal models for robot control," in *2014 IEEE International Symposium on Intelligent Control (ISIC)*, 2014, pp. 1336–1341.

[11] A. K. Pandey and R. Alami, "A framework towards a socially aware mobile robot motion in human-centered dynamic environment," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 5855–5860.

[12] I. Torre, A. Latupeirissa, and C. McGinn, "How context shapes the appropriateness of a robots voice," *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 215–222, 2020.

[13] F. R. L. . F. M. R. . V. Matellan, "Context awareness in shared human-robot environments: Benefits of environment acoustic recognition for user activity classification," *IET Conference Proceedings*, pp. 24 (6 .)–24 (6 .)(1), January 2017.

[14] A. Cosgun and H. I. Christensen, "Context-aware robot navigation using interactively built semantic maps," *Paladyn, Journal of Behavioral Robotics*, vol. 9, no. 1, pp. 254–276, 2018. [Online]. Available: https://doi.org/10.1515/pjbr-2018-0020

[15] H. Liu, Y. Wang, W. Ji, and L. Wang, "A context-aware safety system for human-robot collaboration," *Procedia Manufacturing*, vol. 17, pp. 238–245, 2018, 28th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2018), June 11-14, 2018, Columbus, OH, USAGlobal Integration of Intelligent Manufacturing and Smart Industry for Good of Humanity. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978918311582

[16] Y. Xiao, H. Liang, J. Yuan, and D. Thalmann, *Body Movement Analysis and Recognition*. Springer International Publishing, 09 2016, pp. 31–53.

[17] Z. Zafar, S. H. Paplu, and K. Berns, "Real-time recognition of extroversion-introversion trait in context of human-robot interaction," in *Advances in Service and Industrial Robotics*, N. A. Aspragathos, P. N. Koustoumpardis, and V. C. Moulianitis, Eds. Cham: Springer International Publishing, 2019, pp. 63–70.

[18] Z. Zafar, S. Hussain Paplu, and K. Berns, "Automatic assessment of human personality traits: A step towards intelligent human-robot interaction," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 1–9.

[19] Romotive. (2021) Romo opensource sdk. [Online]. Available: https://github.com/Navideck/Romo-iOS-SDK

[20] iRobot. (2021) Create2 programmable robot by irobot. [Online]. Available: https://store.irobot.com/default/create-programmable-programmable-robot-irobot-create-2/RC65099.html

[21] T. Kohonen, "Self-organized formation of topographically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.

[22] ——, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 2004.

[23] S. Heath, J. Liddle, and J. Wiles, "The challenges of designing a robot for a satisfaction survey: Surveying humans using a social robot," *International Journal of Social Robotics*, vol. 12, pp. 519–533, 2020.

[24] A. M. Althubaiti, "Information bias in health research: definition, pitfalls, and adjustment methods," *Journal of Multidisciplinary Healthcare*, vol. 9, pp. 211 – 217, 2016.

[25] D. Lottridge, M. Chignell, and A. Jovicic, "Affective interaction: Understanding, evaluating, and designing for human emotion," *Reviews of Human Factors and Ergonomics*, vol. 7, no. 1, pp. 197–217, 2011.

14

Paper 4: A minimalistic approach to user-group adaptation of robot behaviors using movement and speech analysis

A minimalistic approach to user-group adaptation of robot behaviors using movement and speech analysis

Morten Roed Frederiksen[1] Kasper Stoy[2]

***Abstract*— Speech characteristics have shown potential as a tool to identify personality factors in humans but usually demands longer interactions or elaborate sensor requirements. This paper presents a novel robot system that uses speech and body movement characteristics to recognize and distinguish between users-groups interacting with it through small sets of interactions. It clusters people with similar characteristics together and measures the affective impact of specific robot behaviors. The system was tested using a custom-created affective robot through 36 interactions with 6 human participants aging from 11 to 70. 186 samples were collected in two different physical contexts and the similarity of the samples for each user was compared. The results indicate that the speech and movement characteristics have the potential as a tool to recognize specific users and as a guide to form user-groups. This was found using only basic sensors available in most robots through a limited set of interactions. The results further highlight that there are significant differences between measurements for the same users in different physical contexts meaning that the participants move and talk differently with each context. This paper suggests combining the speech and movement characteristics with information on the physical context to gain better user adaption in robot behaviors for future projects.**

I. Introduction

When people interact with robots, they tend to anthropomorphize them and change their behavior to fit the interactions, similar to how they would adjust their behaviors when interacting with other humans [1], [2]. This is a particular human skillset that is complex to transfer to a robot. Social robots are often equipped with sensors and software to recognize individual users and attempt to adapt to them. An example of such is Vector by Anki robotics. This robot uses facial recognition to recognize users and subsequently states their names when it sees them to catalyze the forming of a bond between it and the humans it interacts with [3]. Robots that recognize specific users have the potential to adapt their behaviors to fit the preference of them. Mitsuga et al. 2008 investigated adapting robot behaviors to social distances, gaze-meeting-ratio, and gesture coordination and found that the robot could successfully adapt to a single user through an interaction [4].

Dynamically adapting to the current user often requires an elaborate sensor setup to be achieved such as 3d coordinates of the user [5] or motion capture systems [4]. However, some research projects suggest gaining further insight into the current user state using simpler measurement techniques. Bohus et al. 2009 found that using audio, they could recognize and distinguish between the different people and their level of engagement [6]. Mower et al. 2007 used skin temperature, and galvanic skin response to measure the engagement level of the current user [7]. These examples successfully build on the information available in the current context. However, there could be potential opportunities in generalizing on the measured characteristics of the users and optimize the behaviors for the user groups rather than the individuals.

Fig. 1. A test participant interacts with the robot in the larger living room context (6x5m).

Adapting behaviors to specific groups is challenging. There are a near infinite number of different human personality specific characteristics that may influence the adaptation. Personality factors such as extraversion, agreeableness, openness, conscientiousness, and neuroticism found in the "Big Five personality model" introduced by Costa and McCrae 1999 has previously gained attention as usable in human-robot-interaction research projects [8], [9]. Especially the personality factors extraversion and openness to experience are important factors that influence participants' attitude towards robots [10], [11]. Andrist et al. 2015 found that the level of extroversion was an effective personality factor to use as a guide for adapting a robot's behavior [12]. Kaplan et al. 2019 indicated that there was a relation between the level of extroversion in the users and their tendency to anthropomorphize robots, making this a viable candidate personality factor to focus on if the goal is to influence the affective impact of robot behaviors [13]. The level of extroversion is often estimated through personality tests featuring a long series of questions to get a viable read

[1]Morten Roed Frederiksen mrof@itu.dk and [2]Kasper Stoy ksty@itu.dk are affiliated with the REAL lab at the Computer science department of The IT-University of Copenhagen, Rued Langgaards vej 7, 2300 Copenhagen S

on the personality factors. However, speech analysis has shown promising results in establishing personality factors to a certain degree [14].

In this paper, we created and investigated the feasibility of a robot system that groups the users it interacts with by using immediate speech and movement characteristics. Furthermore, the system attempts to achieve this through a minimal set of interactions with the users. With our system we wanted to verify the following hypotheses:

- That the speech characteristics attributes in combination with movement characteristics can sufficiently be used to distinguish a single user and determine the user-group based on the distance between user characteristics. The speech attributes are avg. speech length for a series of questions, average pitch, average answer reaction time, average words per minute, average jitter, shimmer, and average pause duration. (See the following section for the elaboration of these). The movement characteristics are average head movement and average body movement (avg. of any recognized joints).
- That the system can gather sufficient information within a minimal interaction with a participant
- That the attributes can be used to recognize users across multiple different contexts.

The robot system also attempted to verify the affective impact of a set of discrete behaviors on the user-group to adapt its selected level of intensity to the preference of the group. As users with different speech and movement characteristics are encountered, the system attempts to cluster participants sharing similar characteristics. The system was tested by collecting 184 samples of speech and movement characteristics from six participants through 36 interactions in two different physical contexts, using a refined version of our custom-created affective robot “Affecta”. The experiment results indicated the following:

- The selected speech and movement attributes are feasible for identifying users and users-groups.
- The attributes are co-dependant as the results indicate that the variation between users across a single speech or movement attribute is too small to drive the distinction, so a combination of attributes is required.
- The measurements on a single participant vary greatly in different contexts making it difficult to distinguish a participant across multiple contexts.

II. Previous approaches

Assessing personality factors and psychological attributes have previously been used for different purposes in robotics. Rossi et al. 2018 performed reliable screening exams on elderly participants using social robots in psychometric evaluation scenarios [15]. Investigating extroversion as expressed by a robot’s gaze movement was done by Momen et al. 2018. They found that the participants were more positive towards interacting with it when the level of extroversion in the robot matched the level of extroversion in the participants [16]. This aligns with Craenen et al. 2010 that found their participants tended to better like the robot more when they attribute traits to it that came close to their own [17]. Tapus and Mataric 2008 investigated behavior adaption along with the introverted/extroverted space in their rehabilitation robot following personality traits gained from questionnaires following Eysenck Personality Inventory [18], [19]. They matched the robot’s behavior to the level of extroversion as outlined in Eysenck’s “Personality Inventory” and found that the users preferred robots that matched their own level of introversion.

A consistent (but nonsignificant trend) was found by Syrdal et al. 2006 in which the level of extraversion in the participants was associated with the number of uncomfortable directions received from a robot that was toleranced by the user [20]. The same authors in 2009 also tested the Negative Attitude towards Robot Scale, a scale to determine biases towards robots. The motivation for this investigation was to understand how the personality factors in the individual users (plus demographic, technology comfortability, etc.) impact how they view and rate robot behaviors [21]. They found no relation to the evaluation of the behavior of the robot but there was a relation to the participants’ comfort level as they interacted with the robot.

Movement and audio to sample details about the users have also been the focus of previous projects. Nakadai et al. 2003 used motion in conjunction with audio to determine the active speaker in a multi-speaker environment. The performance of the system they made depended on the accuracy of localization. They found that movement motions that were directed at the sound source improve the recognition of the active speaker as it strengthens its ability to separate the speaker from other audio sources [22]. Zafar et al. 2018 used machine learning to classify movement patterns captured by an RGB-D sensor to classify non-verbal extroversion, agreeableness, and neuroticism traits [23]. They used a humanoid robot in different role-play scenarios and reached a high accuracy in the automatic assessment of personality traits.

Batrinca et al. 2012 found that extraversion and Emotional Stability were the easiest factors of the five-factor personality model to automatically detect using video analysis of the participants in different collaborative human-computer-interaction settings [24]. Pianesi et al. 2008 found promising results in classifying extroversion and Locus of Control using audio and video analysis of 1-minute slices of human-to-human interactions [25]. They found that the input from both the participants in the interaction improved the accuracy of their classification.

III. Method

The previous section includes several successful examples of robots adapting to different personality factors of the users. Some of these projects found good results using manual questionnaires, to gain insight into the personality factors of the users, while others successfully used elaborate sensors to record the attributes of the users. However, there may be room for improvements in both simplifying the requirements of the sensors and decreasing the temporal demands of the interaction. If the sensor setup could be less elaborate and the data and temporal demands could be lowered the user adaption could apply to more robot projects.

This study used speech and movement characteristics that were measurable with a relatively simple sensor setup using means available in most robots. The system used microphone and camera sensors as they are often already available on robots for navigation and communication purposes. We measure speech and movement characteristics as they can potentially be accurately measured within a short timespan and through a minimal set of interactions. This paper investigate if this setup is feasible to guide behavior-selection.

Eysenck 1975 detailed a personality inventory with extroversion as a key attribute for different user-groups based on personality factors [18], [26]–[29]. Extroversion can roughly be simplified to how *energetic*, how *talkative*, and how *outgoing* a person is. The attributes measured by the robot in this study were chosen because they relate to this definition and have previously been found viable to indicate the level of extroversion in humans. The definition of extraversion can be translated to some of the measurable characteristics used in this study as follows:

- **Energetic** - We measured: *Head movement*, *Body movement*, *Jitter*, *Shimmer*.
- **Talkative** - We measure: *Words per minute*, *Average answer length*.
- **Outgoing** - We measure: *Average response time*, *Average answer length*.

The level of extroversion in humans has in previous studies been found influental on the human preference in robot behaviors. This means that using similar speech and movement characteristics could be a valid starting point for drive user-group adaption of robot behaviors. The measured attributes are as follows:

- average voice jitter
- average voice shimmer
- average answer length
- average answer reaction time
- average words per minute
- average pause duration

The movement characteristics:

- average head movement
- average body movement

The average voice jitter can be defined as the overall variation in pitch for the user's voice in the interaction. The jitter is recorded as a percentage of the measured fundamental frequency of the voice. The average shimmer can be defined as the overall variation in the volume of the user's voice. The shimmer is recorded in decibel. The average answer length is recorded as the time in seconds measured from the end of the question (when speech recognition starts) until the end of the answer (1.2 seconds after the last word has been recognized). The average answer reaction time is the time in seconds from the end of the question (when the speech recognition starts) until the first word was recognized. Both the estimated answer length and reaction times were dependent on the ability of the speech recognition to recognize the spoken words. The words per minute represent the speaking velocity of the speaker. This estimated value is calculated as the number of words recognized for a single minute based on extending the total time of the total recognized words in the last recorded sentence.

Fig. 2. Affecta V3, The humanoid robot that was used in the interactions with the test participants.

The Affecta robot that was used for the interactions can be seen in Figure 2. The sensors on it used in this paper were proximity sensors, a camera sensor, and a directional microphone. To aid in the accuracy of the recordings of all the audio features, the robot was equipped with an internally powered microphone aimed towards the user in front of the robot. All the measured values were normalized before they were used in comparison with each other. The robot used an iOS app and Apples text-to-speech API to provide some of the speech attributes. The voice jitter, voice shimmer, average pitch, and average pause duration were provided from the Apple Speech analysis API [30]. The iPhone app also utilized the iOS face-tracking and pose-estimation APIs within ARkit to track the individual joint movements of the user [31]. The data gained from the APIs were provided in individual audio frames, and individual video frames. They were all averaged and normalized before use.

The measured values were all normalized and stored on the robot. With each interaction, the robot measured a single sample containing all values averaged over the whole interaction. This sample can be represented as a feature vector containing all the normalized values. This vector holds a single sample of a user's speech and movement characteristics. The distance between two vectors was defined as the sum of squared differences between the individual samples. This distance determines the similarity of two participants' characteristics, and as this value approaches zero, the probability of those two persons being similar increases. If they are close, it is determined that they are part of the same user-group (each group sharing similar speech and movement characteristics).

The Affecta V3 robot used in this paper was equipped with a few preconfigured behaviors. The behavior consisted of simple movement-patterns going backward, forward, and turning from side to side. The behaviors also consisted of simple gestures using head antennas on either side of the robot's head, face animations using the Romo opensource facial animations, and simple audio sounds [32]. The behaviors were nearly identical in their physical movement and gesture patterns but differed in intensity strength with each behavior. Behavior 0 contained no movements or gesturing while behavior 3 contained large and swift movement and gesture patterns. The intensity of each behavior was also evident in its facial animation with behavior 0 being very calm and behavior 3 being highly energetic.

This paper aimed to investigate the feasibility of user-group adaption based on the chosen speech and movement characteristics gathered with a minimal set of interactions. In the experiments, this was approached by letting the robot gather samples through interactions with human participants. In these interactions, the robot also displayed two different behaviors in sequence and obtained the participants' opinions on them. The opinions were obtained by capturing and analyzing images of the participants' facial expressions to classify their current affective status expressed as valence/arousal values. The full experiment details are described in the experimental setup section.

Each question in the interactions amounted to a single sample and more than 40 samples were collected in all from each person. The number varies a bit as some questions had to be repeated when the participants gave an answer that was recorded but not understood by the robot. To establish the validity of the attributes as a tool to distinguish individual users distances between vectors from the same participants and the distances between vectors from different participants were compared. The hypothesis was that the measured vectors will be similar for the samples from the same participant across multiple interactions and different for two different participants.

The participants were sampled through six individual interactions in two different physical contexts to investigate the stability of participant-distinction across different environments. The calculated distance was compared between vectors from the same participants but also with vectors from the same participants across different physical contexts.

IV. Experimental setup

The interaction tests were conducted in two different physical environments, an indoor living room (6x5m) and a smaller indoor bedroom (2x3m). Figure 1 shows a participant interacting with the robot in the larger of the two physical contexts. The data gathered in for this paper was collected as part of a larger experiment on context awarnes and robot behavior-selection. the approach outlined in this section is similar to the experimental setup in Frederiksen and Stoy 2021, although the collected data types are different [33]. The tests included 36 interactions with 6 different test participants aging from 9 to 70. The initial focus of the interactions was to establish the personality traits of the test participants through interactions with the robot. Once established, the interaction continued with the robot measuring the impact of 2 different behaviors by demonstrating them to the user. The test participants interacted individually with the robot in turn, and each interaction lasted for about five minutes. Each participant would interact with the robot three times in each physical context. The participants were informed that the interaction would start as soon as the robot could see them. They were not told to perform any actions other than to interact with the robot and to answer any questions the robot would ask them. The following steps were performed with each participant:

1. When the robot would detect the face of a participant, it would start the interaction by asking the participant the following: *"Hello there, can I ask you a question?"*.
2. If the participant agreed, the robot would continue with: *"Could I ask you to tell me a bit about yourself, like who are you and what do you do?"*.
3. The robot would listen to the participantss answer and continue: *"Okay.. thank you, that sounds nice. Listen I actually need your help.. I would like you to assess my behavior.. That's really really difficult for me and I would appreciate your honesty. I will show you two different behaviors and I would like you to compare the two. Are you ready? Here is the first.."*.
4. The robot would select two random behaviors and start by demonstrating the first to the participant.
5. Before using the behavior, the robot would say: *"Okay.. watch carefully.. here comes the second behavior"*. After which the robot showed the second behavior.
6. The robot recorded the participant's reaction to the behavior and would try to classify his or her current affective status expressed as a normalized value for valence and arousal levels.

The robot would track the participant's speech- and movement characteristics with each answer received from the participant. The interaction contained a step where the robot could accept or refuse to continue and a step where it needed to understand the preference of the user. If the robot did not understand the answer in one of these steps (1,2) it would repeat the question. The robot demonstrated two behaviors in each interaction. Immediately following the demonstration, the robot would attempt to track the participant's face and record an immediate reaction to the demonstrated behavior. The robot would ensure the visibility of the participants face before the demonstration by asking the participant *"Where did you go? I can't see your face anymore. Can you please move so I can see your face?"*. The captured image sequence of the user's face was analyzed with each demonstrated behavior using a pre-trained convolutional neural network (trained on annotated data from the AffectNet dataset) to classify the participant's current affective status as expressed in Russel 1980 with valence- and arousal levels [34], [35]. The detected pleasantness level for each behavior would determine the users preference in the different behaviors.

Group	answer length	body move-ment	head move-ment	reaction time	voice jitter	voice shim-mer	pause dura-tion
All	2,754	0,021	0,088	0,536	3,442	1,710	0,269
Single	1,235	0,009	0,049	0,159	1,544	0,517	0,202
Ratio	0,449	0,423	0,559	0,297	0,448	0,302	0,752

TABLE I
THE RESULTING STANDARD DEVIATIONS FOR EACH MEASURED ATTRIBUTE. THE TOP ROW CONTAINS AVERAGES AND STANDARD DEV. ACROSS ALL RECORDED SAMPLES, WHILE THE BLUE ROWS SHOW AVERAGES AND STANDARD DEV. AS AN AVERAGE FOR EACH USER. THE GREEN ROWS SHOW THE RATIO BETWEEN THE SPREAD OF DATA FOR ALL SAMPLES AND SAMPLES AVERAGED PER INDIVIDUAL PARTICIPANT.

V. RESULTS

With the results we are interested in seing how close each sample of speech and movement characteristics lies for each participant and in general. Table I shows the standard deviations for the collected samples across all participants and the standard deviation of each individual participant's samples. The values are calculated from the raw recorded numbers and have not been normalized. The individual measurements were first calculated for the individual users and the averages of those numbers were recorded. The results highlights that on average the samples for a single participant lie closer than the average of samples in general. This is evident in the ratio between the standard deviations for all participants and per individual participants with numbers ranging from 0.29 to 0.75. All ratios for the standard deviations were less than one. With the exception of the standard deviation for the average pause duration (0.75), and average head movement (0.55) all attributes had a ratio below 0.5.

For the samples gathered in the different contexts the distance between them reveals the how well the speech and movement attributes can be used to distinguish users in different physical locations. Figure 3 shows all six participants and the Euclidean distance between the vectors containing the average of their respective samples in each context. Each participant was tested in two contexts (circles with similar colors) and as the visualization highlights, the distances between the average measurements are further between contexts than between individual participants in each context. The calculated average distances between measurements in each context and between the two contexts are: 0.30 for the large (5x6m) context, 0.23 for the smaller (2x3m) context, and 0.54 for the measurements in the one context compared to the measurements from the other. Based on the measurements, these numbers show that the vector of personal traits of participants are more similar to each other in each physical context than similar to themselves in a different context.

VI. APPLYING THE FINDINGS

The results indicate that the chosen speech- and movement characteristics can adequately be used to identify users, even when gathered through a minimal set of interactions. This

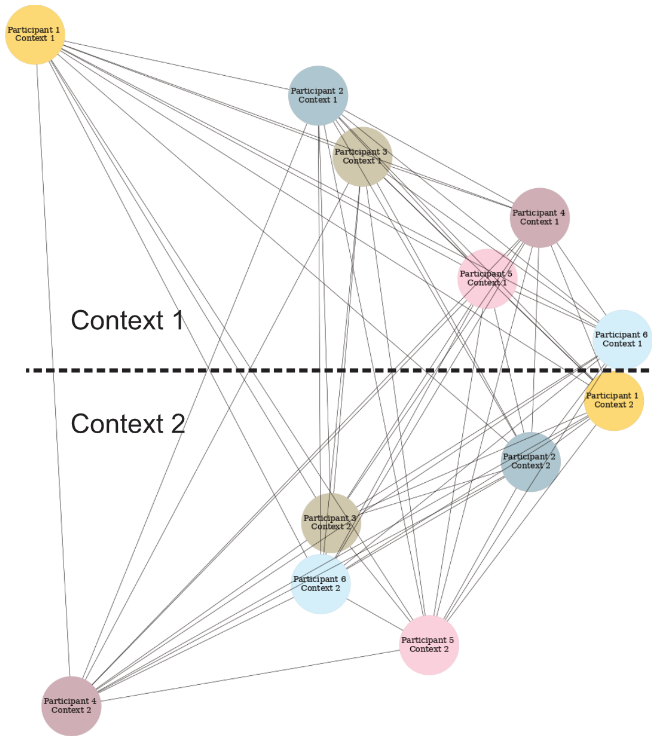

Fig. 3. The participants and the distance relationship between the average of their respective samples in each context. The colored circles represents participants and the edges between them is the Euclidean distance between the two participants' samples. Each participant was tested in two contexts and the identical colors of the circles represent the same participant in different contexts.

is evident in the gathered samples as the average distances between the samples gathered from a single participant are smaller than the average distance between the samples from two different participants. The measurements for a single attribute may be too close for it to provide enough detail to distinguish participants based on that alone. A combination of points is needed but a subset of the ones we use may be sufficient. We did not investigate this further.

The distance between samples collected from the same participants but in different physical contexts is higher than the average distance between different users in the same context. This poses a challenge for the recognition of users and behavior adaption across multiple physical contexts. It also indicates that the participants potentially alter the way they talk and move with each context and the robot may register that. This may indicate that any robot system should not rely on a single measurement when trying to establish user-groups based on personality factors. It is a continuous process and the behavior patterns humans exhibit in one context may be different in the next. Also, this could indicate that for a social robot to appear more human-like it could benefit from adapting its different behaviors to not only the personality factors of the humans it interacts with but also the physical context as well. It would make sense to combine the knowledge gained from the speech and movement measurements with additional info on the physical context. E.g. a single distance measurement of the robot's current physical surroundings may provide enough additional information to gain a better adaption to its users than relying solely on data gained from the interaction.

This study also investigated whether the speech and movement attributes could be used to drive robot behavior adaption to a specific user group rather than individual persons. If this was the case, it should be possible to generalize over the behavior preference from participants sharing similar characteristics. This would mean that there could be established a correlation between the measured characteristics and the participants' preferences in the different robot behaviors. There were four behaviors in all and they differed in the amount of movement the robot used and how energetic the robot acted. The speech and movement characteristics we gathered in the experiments related to the level of extroversion in the participants. As extroversion has previously been shown to influence humans' preference in robot behaviors, a relation between the detected user-group characteristics and their behavior preferences seems feasible [12], [19], [36]. However, the data we gathered in these experiments were insufficient to support anything significant about such a relationship. The problem may arise from the difficulty in establishing a personality type using data measured in a limited interaction using three simple questions and movement patterns. The Eysenck personality inventory test uses 50+ questions and likewise suffers from being unprecise.

The distance between speech and movement attribute vectors was calculated with equal emphasis on all attributes. It could potentially improve the correlation between attribute vector distances and behavior preference vector differences if each attribute was adjusted by a weight to make it more or less influential on the vector distance. This could be handled by post-processing the data and adjusting the weights based on the degree of mismatch between the two calculated distances. A potential pitfall with such a process is that the data might be overfitted to a specific dataset making it less usable as future samples are added.

Overall the system for user-group detection is a simplification of some of the psychological processes that occur in all of us as we interact with other people. It will often be wrong as it is working on limited input and limited temporal constraints. However, the same impreciseness is evident in most humans as we estimate the personality of other humans. This is acceptable as we don't have to determine the definitive correct behavior for a given interaction. It is enough to get a ballpark estimation of other people to get a better starting point for a successful interaction. The same can be said about the robot. The limited amount of time spent in the interaction and simple sensor setup may decrease the accuracy of how well the robot detects different user-groups. A fair degree of uncertainty can be tolerated in its estimation of user-groups and behavior adaption to them. Even the lightest information can lead it to a better starting point for successful communication with its users. The user-group adaption system is also not necessarily meant to work in unison but could benefit from being combined with other contextual information systems to provide a more detailed image of the user and the current context.

VII. Conclusion

This paper aimed to investigate the feasibility of using speech and movement characteristics, measured over a short interaction time, to sufficiently detect individual users and investigate whether it works across multiple different physical contexts. The hypothesis has been tested using a custom-built robot in 72 interactions with six participants across two different physical contexts. It was found that the speech and movement characteristics were sufficient to distinguish a participant in a single context. Defining user-groups on the distance between user-characteristics is feasible. It was also found that the attributes are co-dependant as the results indicate that the variation between users across a single speech or movement attribute is too small to drive the distinction, so a combination of attributes is required. The measurements on a single participant vary greatly in different contexts making it difficult to distinguish a participant across multiple contexts. The project showed that measuring aspects of the users' personality factors is possible with some tolerated impreciseness. The authors suggest combining the measured data with additional contextual information systems to better guide user-group adaption in future robot projects.

References

[1] S. Fussell, S. Kiesler, L. Setlock, and V. Yew, "How people anthropomorphize robots," in *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, 01 2008, pp. 145–152.

[2] B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people and pla," *Bibliovault OAI Repository, the University of Chicago Press*, 01 1996.

[3] Anki, "Vector: The home robot with interactive ai technology," 06-20-2020. [Online]. Available: www.anki.com/en-us/vector

[4] N. Mitsunaga, C. Smith, T. Kanda, H. Ishiguro, and N. Hagita, "Adapting robot behavior for human–robot interaction," *Robotics, IEEE Transactions on*, vol. 24, pp. 911 – 916, 09 2008.

[5] A. Tapus and A. Amir, "User adaptable robot behavior," *Proceedings of the 2011 International Conference on Collaboration Technologies and Systems, CTS 2011*, 05 2011.

[6] D. Bohus and E. Horvitz, "Learning to predict engagement with a spoken dialog system in open-world settings," in *Proceedings of the SIGDIAL 2009 Conference*, 2009, pp. 244–252.

[7] E. Mower, D. J. Feil-Seifer, M. J. Mataric, and S. Narayanan, "Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements," in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 2007, pp. 1125–1130.

[8] P. Costa and R. McCrae, "A five-factor theory of personality," *The Five-Factor Model of Personality: Theoretical Perspectives*, vol. 2, pp. 51–87, 01 1999.

[9] L. P. R. Jr., R. Alahmad, C. Esterwood, S. Kim, S. You, and Q. Zhang, "A review of personality in humanrobot interactions," *Foundations and Trends in Information Systems*, vol. 4, no. 2, pp. 107–212, 2020.

[10] U. Morsunbul, "Human-robot interaction : How do personality traits affect attitudes towards robot ?" in *Journal of Human sciences*, vol. 16. No 2, 2019.

[11] S. L. Müller and A. Richert, "The big-five personality dimensions and attitudes to-wards robots: A cross sectional study," in *PETRA '18*, 2018.

[12] S. Andrist, B. Mutlu, and A. Tapus, *Look Like Me: Matching Robot Personality via Gaze to Increase Motivation*. New York, NY, USA: Association for Computing Machinery, 2015, p. 36033612. [Online]. Available: https://doi.org/10.1145/2702123.2702592

[13] A. D. Kaplan, T. Sanders, and P. A. Hancock, "The relationship between extroversion and the tendency to anthropomorphize robots: A bayesian analysis," *Frontiers in Robotics and AI*, vol. 5, p. 135, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2018.00135

[14] A. Guidi, C. Gentili, E. Scilingo, and N. Vanello, "Analysis of speech features and personality traits," *Biomedical Signal Processing and Control*, vol. 51, pp. 1–7, 05 2019.

[15] S. Rossi, G. Santangelo, M. Staffa, S. Varrasi, D. Conti, and A. G. D. Nuovo, "Psychometric evaluation supported by a social robot: Personality factors and technology acceptance," *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 802–807, 2018.

[16] A. Momen and E. Wiese, "Noticing extroversion effects attention: How robot and participant personality affect gaze cueing," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 62, no. 1, pp. 1557–1561, 2018. [Online]. Available: https://doi.org/10.1177/1541931218621352

[17] B. G. W. Craenen, A. A. Deshmukh, M. E. Foster, and A. Vinciarelli, "Shaping gestures to shape personality: Big-five traits, godspeed scores and the similarity-attraction effect," in *AAMAS*, 2018.

[18] H. Eysenck, "The structure of human personality," in ., 1953.

[19] A. Tapus and M. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance." in *AAAI Spring Symposium - Technical Report*, 01 2008, pp. 133–140.

[20] D. S. Syrdal, K. Dautenhahn, S. Woods, M. L. Walters, and K. L. Koay, "'doing the right thing wrong' - personality and tolerance to uncomfortable robot approaches," in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006, pp. 183–188.

[21] D. S. SYRDAL, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," *Proc. New Frontiers in HRI Symposium, Edinburgh, 2009*, pp. 109–115, 2009. [Online]. Available: https://ci.nii.ac.jp/naid/20000716587/en/

[22] K. Nakadai, H. Okuno, and H. Kitano, "Robot recognizes three simultaneous speech by active audition," *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, vol. 1, pp. 398–405 vol.1, 2003.

[23] Z. Zafar, S. H. Paplu, and K. Berns, "Automatic assessment of human personality traits: A step towards intelligent human-robot interaction," *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 1–9, 2018.

[24] L. Batrinca, B. Lepri, N. Mana, and F. Pianesi, "Multimodal recognition of personality traits in human-computer collaborative tasks," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ser. ICMI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 3946.

[25] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro, "Multimodal recognition of personality traits in social interactions," in *Proceedings of ICMI 2008*, 01 2008, pp. 53–60.

[26] H. J. Eysenck and S. B. G. Eysenck, "The eysenck personality inventory," in ., 1965.

[27] A. M. Bodling and T. Martin, *Eysenck Personality Inventory*. New York, NY: Springer New York, 2011, pp. 1007–1008.

[28] H. J. Eysenck and S. B. G. Eysenck, "Manual of the eysenck personality questionnaire," in ., 1975.

[29] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications." *Journal of personality*, vol. 60 2, pp. 175–215, 1992.

[30] Apple. (2021) Apple ios speech framework. [Online]. Available: https://developer.apple.com/documentation/speech

[31] ——. (2021) Apple ios arkit framework. [Online]. Available: https://developer.apple.com/augmented-reality/

[32] Romotive. (2021) Romo opensource sdk. [Online]. Available: https://github.com/Navideck/Romo-iOS-SDK

[33] M. R. Frederiksen and K. Stoy, "Adaptable context-based behavior selection in autonomous robots," in *Submitted to 2021 IEEE/RSJ International Conference On Intelligent Robots and Systems*, 01 2021.

[34] A. Mollahosseini, B. Hasani, and M. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. PP, 08 2017.

[35] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 12 1980.

[36] A. Tapus and M. Mataric, "User personality matching with a hands-off robot for post-stroke rehabilitation therapy," in *Springer Tracts in Advanced Robotics*, vol. 39, 01 2006, pp. 165–175.

15 Bibliography

[1] A. Bourai, T. Baltrusaitis, and L.-P. Morency, "Automatically predicting human knowledgeability through non-verbal cues," *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017.

[2] T. Polzin, "Verbal and non-verbal cues in the communication of emotions," *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, vol. 4, pp. 2429–2432 vol.4, 2000.

[3] J. Vroon, G. Englebienne, and V. Evers, "Detecting perceived appropriateness of a robot's social positioning behavior from non-verbal cues," *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*, pp. 216–225, 2019.

[4] V. Braitenberg, *Vehicles, Experiments in Synthetic Psychology.* MIT Press, 1986.

[5] R. Brooks, "Integrated systems based on behaviors," *SIGART Bull.*, vol. 2, pp. 46–50, 1991.

[6] P. Agre, "Lucy a. suchman, plans and situated actions: The problem of human-machine commuinication (cambridge university press, cambridge 1987)," *Artif. Intell.*, vol. 43, pp. 369–384, 1990.

[7] J. Verhagen, R. Berghe, O. Oudgenoeg-Paz, A. Küntay, and P. Leseman, "Children's reliance on the non-verbal cues of a robot versus a human," *PLoS ONE*, vol. 14, 2019.

[8] S. Fussell, S. Kiesler, L. Setlock, and V. Yew, "How people anthropomorphize robots," in *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots*, 01 2008, pp. 145–152.

[9] M. R. Frederiksen and K. Støy, "A systematic comparison of affective robot expression modalities," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1385–1392, 2019.

[10] T. Numanoglu, E. Erzin, Y. Yemez, and M. Sezgin, "Head nod detection in dyadic conversations," *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, 2019.

[11] G. Dezecache, H. Mercier, and T. C. Scott-Phillips, "An evolutionary approach to emotional communication," *Journal of Pragmatics*, vol. 59, pp. 221–233, 2013.

[12] D. E. Karreman, B. Dijk, and V. Evers, "Contextual analysis of human non-verbal guide behaviors to inform the development of frog, the fun robotic outdoor guide," in *HBU*, 2012.

[13] A. Anikin, R. Bååth, and T. Persson, "Human non-linguistic vocal repertoire: Call types and their meaning," *Journal of Nonverbal Behavior*, vol. 42, pp. 53 – 80, 2018.

[14] R. Pfeifer and J. Bongard, *How the Body Shapes the Way We Think: a New View of Intelligence*. The MIT Press, 01 2007.

[15] F. Gee, W. Browne, and K. Kawamura, "Uncanny valley revisited," *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pp. 151–157, 2005.

[16] A. Barco, C. D. Jong, J. Peter, R. Kühne, and C. L. V. Straten, "Robot morphology and children's perception of social robots: An exploratory study," *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020.

[17] R. Pfeifer, "Building "fungus eaters": Design principles of autonomous agents," *From Animals to Animats*, 09 1996.

[18] P. Menezes, J. Quintas, and J. Dias, "The role of context information in human-robot interaction," in *Proceedings of 23rd IEEE International Symposium on Robot And Human Interactive Communication (IEEE RO-MAN 2014)*, 2014.

[19] D. Dennett, "Cognitive wheels: The frame problem of ai," in *The Philosophy of Artificial Intelligence*, 1990.

[20] M. Nakajima, L. I. Schmitt, and M. M. Halassa, "Prefrontal cortex regulates sensory filtering through a basal ganglia-to-thalamus pathway," *Neuron*, vol. 103, pp. 445–458.e10, 2019.

[21] C. Cordes, "The human adaptation for culture and its behavioral implications," *Journal of Bioeconomics*, vol. 6, pp. 143–163, 2004.

[22] S. Andrist, B. Mutlu, and A. Tapus, "Look like me: Matching robot personality via gaze to increase motivation," *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015.

[23] B. Soltani and M. Keyvanara, "Cultural intelligence and social adaptability: A comparison between iranian and non-iranian dormitory students of isfahan university of medical sciences," *Materia Socio-Medica*, vol. 25, pp. 40 – 43, 2013.

[24] 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, pp. 457–476, 2013.

[25] M. Khoramshahi and A. Billard, "A dynamical system approach to task-adaptation in physical human–robot interaction," *Autonomous Robots*, vol. 43, pp. 927–946, 2019.

[26] R. Pfeifer, M. Lungarella, and F. Iida, "Self-organization, embodiment, and biologically inspired robotics," *Science*, vol. 318, no. 5853, pp. 1088–1093, 2007.

[27] D. Ullrich, "Robot personality insights. designing suitable robot personalities for different domains," *i-com*, vol. 16, pp. 57 – 67, 2017.

[28] N. Epley, A. Waytz, and J. Cacioppo, "On seeing human: a three-factor theory of anthropomorphism." *Psychological review*, vol. 114 4, pp. 864–86, 2007.

[29] F. Heider and M. Simmel, "An experimental study of apparent behavior," *American Journal of Psychology*, vol. 57, pp. 243–259, 1944.

[30] O. Holland, "Exploration and high adventure: The legacy of grey walter," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 361, pp. 2085–121, 11 2003.

[31] W. Walter, "An imitation of life," *Scientific American*, vol. 182, pp. 42–45, 1950.

[32] R. Brooks, "A robot that walks; emergent behaviors from a carefully evolved network," *Proceedings, 1989 International Conference on Robotics and Automation*, pp. 692–4+2 vol.2, 1989.

[33] ——, "Elephants don't play chess," in *Robotics Auton. Syst.*, vol. 6, 1990, pp. 3–15.

[34] V. V. Hafner, H. Kunz, and R. Pfeifer, "An investigation into obstacle avoidance as an 'emergent'behaviour from two different perspectives," in *Proceedings of the EPSRC/BBSRC International Workshop on Biologically-Inspired Robotics: The Legacy of W. Grey Walter*, 2002, pp. 166–173.

[35] S. Bovet and R. Pfeifer, "Emergence of coherent behaviors from homogenous sensorimotor coupling," in *ICAR'05. Proceedings., 12th International Conference on Advanced Robotics, 2005.* IEEE, 2005, pp. 324–330.

[36] M. R. Frederiksen and K. Støy, "Robots can defuse high-intensity conflict situations," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11 376–11 382, 2020.

[37] M. J. Mataric, "Behavioral synergy without explicit integration," *SIGART Bull.*, vol. 2, no. 4, p. 130–133, Jul. 1991. [Online]. Available: https://doi.org/10.1145/122344.122370

[38] M. Mataric and F. Michaud, *Behavior-Based Systems.* Springer, 01 2008, pp. 891–909.

[39] C. Breazeal and B. Scassellati, "A context-dependent attention system for a social robot," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2, 08 1999.

[40] R. A. Brooks, C. Breazeal, M. Marjanović, B. Scassellati, and M. M. Williamson, "The cog project: Building a humanoid robot," in *Computation for Metaphors, Analogy, and Agents*, C. L. Nehaniv, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 52–87.

[41] C. Breazeal, J. Gray, and M. Berin, "Mindreading as a foundational skill for socially intelligent robots," in *Robotics Research*, M. Kaneko and Y. Nakamura, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 383–394.

[42] A. Thomaz, M. Berlin, and C. Breazeal, "An embodied computational model of social referencing," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 2005, pp. 591–598.

[43] M. R. Frederiksen and K. Støy, "On the causality between affective impact and coordinated human-robot reactions," *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 488–494, 2020.

[44] M. R. Frederiksen and K. Stoy, "Adaptable context-based behavior selection in autonomous robots," in *Currently under review: 2021 IEEE/RSJ International Conference On Intelligent Robots and Systems (IROS)*, 01 2021.

[45] ——, "A minimalistic approach to user-group adaptation of robot behaviors using movement and speech analysis," in *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).*, 01 2021.

[46] M. R. Frederiksen and K. Støy, "Augmenting the audio-based expression modality of a non-affective robot," *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 144–149, 2019.

[47] W. J. Freeman, "W.g. walter: The living brain," in *Brain Theory*, G. Palm and A. Aertsen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 237–238.

[48] G. M. Levitt, *The Turk, Chess Automation.* McFarland and Company, Inc. Publishers, 2000.

[49] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of oz studies: Why and how," in *Proceedings of the 1st International Conference on Intelligent User Interfaces*, ser. IUI '93. New York, NY, USA: Association for Computing Machinery, 1993, p. 193–200. [Online]. Available: https://doi.org/10.1145/169891.169968

[50] K. Čapek, *R.U.R. (Rossum's Universal Robots) 1890-1938.* New York :Penguin Books, 2004.

[51] N. Nilsson, "Shakey the robot, artificial intelligence center, standford robotics institute," 1984.

[52] R. Brooks, "A robust layered control system for a mobile robot," *IEEE J. Robotics Autom.*, vol. 2, pp. 14–23, 1986.

[53] Y. Wei and W. Liu, "Autonomous and rational behavior of sony aibo robot," 2005.

[54] (2018, Nov.) Pepper by softbank robotics. [Online]. Available: https://www.softbankrobotics.com/emea/en/robots/pepper

[55] B. D. Carolis, V. Carofiglio, I. Grimandli, N. Macchiarulo, G. Palestra, and O. Pino, "Using the pepper robot in cognitive stimulation therapy for people with mild cognitive impairment and mild dementia," in *ACHI 2020 : The Thirteenth International Conference on Advances in Computer-Human Interactions*, 2020.

[56] (2011, Nov.) Nao by softbank robotics. [Online]. Available: https://www.softbankrobotics.com/emea/en/robots/pepper

[57] R. Andreasson, B. Alenljung, E. Billing, and R. J. Lowe, "Affective touch in human–robot interaction: Conveying emotion to the nao robot," *International Journal of Social Robotics*, vol. 10, pp. 473–491, 2018.

[58] E. Frid, R. Bresin, and S. Alexanderson, "Perception of mechanical sounds inherent to expressive gestures of a nao robot - implications for movement sonification of humanoids," in *Proceedings of the 15th Sound and Music Computing Conference*, 2018.

[59] O. Tutsoy, D. E. Barkana, and S. Colak, "Learning to balance an nao robot using reinforcement learning with symbolic inverse kinematic," *Transactions of the Institute of Measurement and Control*, vol. 39, pp. 1735 – 1748, 2017.

[60] T. Shibata, K. Inoue, and R. Irie, "Emotional robot for intelligent system-artificial emotional creature project," in *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, 1996, pp. 466–471.

[61] (2018, Nov.) Paro by paro robots. [Online]. Available: http://www.parorobots.com/

[62] (2021, Nov.) Sophia by hansen robotics. [Online]. Available: https://www.hansonrobotics.com

[63] S. Park, H. Lee, D. Hanson, and P. Oh, "Sophia-hubo's arm motion generation for a handshake and gestures," *2018 15th International Conference on Ubiquitous Robots (UR)*, pp. 511–515, 2018.

[64] (2019, Nov.) The 28th ieee international conference on robot and human interactive communication. [Online]. Available: https://ro-man2019.org/sophia-the-robot/

[65] A. V. D. Pütten, F. P. Schulte, S. Eimler, S. Sobieraj, L. Hoffmann, S. Maderwald, M. Brand, and N. Krämer, "Investigations on empathy towards humans and robots using fmri," *Comput. Hum. Behav.*, vol. 33, pp. 201–212, 2014.

[66] J. Zamfirescu-Pereira, D. Sirkin, D. Goedicke, R. LC, N. Friedman, I. Mandel, N. Martelaro, and W. Ju, "Fake it to make it: Exploratory prototyping in hri," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 19–28. [Online]. Available: https://doi.org/10.1145/3434074.3446909

[67] H. Erel, Y. Cohen, K. Shafrir, S. D. Levy, I. D. Vidra, T. Shem Tov, and O. Zuckerman, "Excluded by robots: Can robot-robot-human interaction lead to ostracism?" in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 312–321. [Online]. Available: https://doi.org/10.1145/3434073.3444648

[68] P. R. Bana, Y.-L. Tsai, and H. Knight, "Sanitizerbot: A hand sanitizer service robot," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 661. [Online]. Available: https://doi.org/10.1145/3434074.3447274

[69] Y.-L. Tsai, P. Bana, and H. Knight, "Can service robots help best practices for covid?" *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021.

[70] R. Simmons and H. Knight, "Keep on dancing: Effects of expressive motion mimicry," *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 720–727, 2017.

[71] C. Herrera Perez and E. I. Barakova, *Expressivity Comes First, Movement Follows: Embodied Interaction as Intrinsically Expressive Driver of Robot Behaviour.* Cham: Springer International Publishing, 2020, pp. 299–313.

[72] M. Sharma, D. Hildebrandt, G. Newman, J. Young, and M. R. Eskicioglu, "Communicating affect via flight path exploring use of the laban effort system for designing affective locomotion paths," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 293–300, 2013.

[73] A. Bevins and B. A. Duncan, "Aerial flight paths for communication: How participants perceive and intend to respond to drone movements," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 16–23. [Online]. Available: https://doi.org/10.1145/3434073.3444645

[74] H. Knight and R. Simmons, "Expressive motion with x, y and theta: Laban effort features for mobile robots," *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pp. 267–273, 2014.

[75] A. Langer and S. Levy-Tzedek, *Priming and Timing in Human-Robot Interactions.* Cham: Springer International Publishing, 2020, pp. 335–350.

[76] R. Mead, A. Atrash, and M. Mataric, "Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction," *International Journal of Social Robotics*, vol. 5, pp. 367–378, 2013.

[77] R. Kitagawa, Y. Liu, and T. Kanda, "Human-inspired motion planning for omni-directional social robots," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 34–42. [Online]. Available: https://doi.org/10.1145/3434073.3444679

[78] M. Saerbeck and C. Bartneck, "Perception of affect elicited by robot motion," *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 53–60, 2010.

[79] G. Yoshioka, T. Sakamoto, and Y. Takeuchi, "Inferring affective states from observation of a robot's simple movements," *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 185–190, 2015.

[80] H. Knight, M. Veloso, and R. Simmons, “Taking candy from a robot: Speed features and candy accessibility predict human response,” *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 355–362, 2015.

[81] J. M. A. Fernandez and A. Bonarini, “Enriching robot’s actions with affective movements,” *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 424–429, 2017.

[82] C. L. Bethel, K. Salomon, and R. R. Murphy, “Preliminary results: Humans find emotive non-anthropomorphic robots more calming,” *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 291–292, 2009.

[83] C. O’Brien, M. O’Mara, J. Issartel, and C. McGinn, “Exploring the design space of therapeutic robot companions for children,” in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 243–251. [Online]. Available: https://doi.org/10.1145/3434073.3444669

[84] E. Cha and M. Mataric, “Using nonverbal signals to request help during human-robot collaboration,” *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5070–5076, 2016.

[85] S. Müller, C. Schröter, and H. Groß, “Smart fur tactile sensor for a socially assistive mobile robot,” in *ICIRA*, 2015.

[86] H. J. Schellin, T. N. Oberley, K. M. Patterson, B. young Kim, K. Haring, C. C. Tossell, E. Phillips, and E. Visser, “Man’s new best friend? strengthening human-robot dog bonding by enhancing the doglikeness of sony’s aibo,” *2020 Systems and Information Engineering Design Symposium (SIEDS)*, pp. 1–6, 2020.

[87] M. Fujita, “Aibo: Toward the era of digital creatures,” *I. J. Robotics Res.*, vol. 20, pp. 781–794, 2001.

[88] W. D. Stiehl, C. Breazeal, K.-H. Han, J. Lieberman, L. Lalla, A. Z. Maymin, J. Salinas, D. Fuentes, R. L. Toscano, C. H. Tong, A. Kishore, M. Berlin, and J. Gray, “The huggable: a therapeutic robotic companion for relational, affective touch,” *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.*, vol. 2, pp. 1290–1291, 2006.

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

[90] L. Boccanfuso, E. S. Kim, J. C. Snider, Q. Wang, C. A. Wall, L. DiNicola, G. Greco, F. Shic, B. Scassellati, L. Flink, S. Lansiquot, K. Chawarska, and P. Ventola, "Autonomously detecting interaction with an affective robot to explore connection to developmental ability," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7, 2015.

[91] A. Singh and J. E. Young, "A dog tail for communicating robotic states," *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 417–417, 2013.

[92] A. E. Block, S. Christen, R. Gassert, O. Hilliges, and K. J. Kuchenbecker, "The six hug commandments: Design and evaluation of a human-sized hugging robot with visual and haptic perception," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 380–388. [Online]. Available: https://doi.org/10.1145/3434073.3444656

[93] H. Knight, T. Lee, B. Hallawell, and W. Ju, "I get it already! the influence of chairbot motion gestures on bystander response," *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 443–448, 2017.

[94] M. Panteris, S. Manschitz, and S. Calinon, "Learning, generating and adapting wave gestures for expressive human-robot interaction," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 386–388. [Online]. Available: https://doi.org/10.1145/3371382.3378286

[95] I. Cohen, R. Looije, and M. A. Neerincx, "Child's recognition of emotions in robot's face and body," *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 123–124, 2011.

[96] J. Xu, J. Broekens, K. V. Hindriks, and M. A. Neerincx, "Effects of a robotic storyteller's moody gestures on storytelling perception," *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 449–455, 2015.

[97] H. Knight and M. Gray, "Acting lesson with robot: Emotional gestures," *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 407–407, 2012.

[98] L. Rincon, E. Coronado, H. Hendra, J. Phan, Z. Zainalkefli, and G. Venture, "Expressive states with a robot arm using adaptive fuzzy and robust predictive controllers," *2018 3rd International Conference on Control and Robotics Engineering (ICCRE)*, pp. 11–15, 2018.

[99] A. Mehrabian and J. Russell, "An approach to environment psychology," *Cambridge, MA*, 01 1974.

[100] H. Lee, Y. Kim, and M. Kim, "Come on in!: A strategic way to intend approachability to a space by using motions of a robotic partition," *2013 IEEE RO-MAN*, pp. 441–446, 2013.

[101] C. Yu and A. Tapus, "Srg3: Speech-driven robot gesture generation with gan," *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 759–766, 2020.

[102] F. A. Robinson, M. Velonaki, and O. Bown, "Smooth operator: Tuning robot perception through artificial movement sound," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 53–62. [Online]. Available: https://doi.org/10.1145/3434073.3444658

[103] S. Rossi, M. Larafa, and M. Ruocco, "Emotional and behavioural distraction by a social robot for children anxiety reduction during vaccination," *International Journal of Social Robotics*, vol. 12, pp. 765–777, 2020.

[104] A. Pipitone and A. Chella, "What robots want? hearing the inner voice of a robot," *iScience*, vol. 24, 2021.

[105] K. Winkle and P. Bremner, "Investigating the real world impact of emotion portrayal through robot voice and motion," *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 627–634, 2017.

[106] C. Becker-Asano and H. Ishiguro, "Laughter in social robotics–no laughing matter," *Intl. Workshop on Social Intelligence Design (SID2009)*, 01 2009.

[107] J. H. Lui, H. Samani, and K. yu Tien, "An affective mood booster robot based on emotional processing unit," *2017 International Automatic Control Conference (CACS)*, pp. 1–6, 2017.

[108] K. Darling, "'who's johnny?' anthropomorphic framing in human-robot interaction, integration, and policy," *SSRN Electronic Journal*, 01 2015.

[109] K. Zawieska, B. Duffy, and A. Sprońska, "Understanding anthropomorphisation in social robotics," *Pomiary Automatyka Robotyka*, pp. 78–82, 11 2012.

[110] M. Natarajan and M. Gombolay, "Effects of anthropomorphism and accountability on trust in human robot interaction," *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020.

[111] Q. R. M. Hover, E. Velner, T. Beelen, M. Boon, and K. P. Truong, "Uncanny, sexy, and threatening robots: The online community's attitude to and perceptions of robots varying in humanlikeness and gender," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 119–128. [Online]. Available: https://doi.org/10.1145/3434073.3444661

[112] S. Marchesi, N. Spatola, J. Perez-Osorio, and A. Wykowska, "Human vs humanoid. a behavioral investigation of the individual tendency to adopt the intentional stance," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 332–340. [Online]. Available: https://doi.org/10.1145/3434073.3444663

[113] C. W. Becker-Asano and H. Ishiguro, "Evaluating facial displays of emotion for the android robot geminoid f," *2011 IEEE Workshop on Affective Computational Intelligence (WACI)*, pp. 1–8, 2011.

[114] Z. Faraj, M. Selamet, C. Morales, P. Torres, M. Hossain, B. Chen, and H. Lipson, "Facially expressive humanoid robotic face," *HardwareX*, vol. 9, p. e00117, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468067220300262

[115] F. Hegel, S. Gieselmann, A. Peters, P. Holthaus, and B. Wrede, "Towards a typology of meaningful signals and cues in social robotics," *2011 RO-MAN*, pp. 72–78, 2011.

[116] E. C. Collins, T. Prescott, and B. Mitchinson, "Saying it with light: A pilot study of affective communication using the miro robot," in *Living Machines*, 2015.

[117] C. Breazeal, A. G. Brooks, J. Gray, G. Hoffman, C. D. Kidd, H. Lee, J. Lieberman, A. L. Thomaz, and D. Chilongo, "Tutelage and collaboration for humanoid robots," *I. J. Humanoid Robotics*, vol. 1, pp. 315–348, 2004.

[118] L. Cañamero and M. Lewis, "Making new "new ai" friends: Designing a social robot for diabetic children from an embodied ai perspective," *International Journal of Social Robotics*, vol. 8, pp. 523–537, 2016.

[119] N. Mirnig, M. Giuliani, G. Stollnberger, S. Stadler, R. Buchner, and M. Tscheligi, "Impact of robot actions on social signals and reaction times in hri error situations," in *ICSR*, 2015.

[120] D. Bohus and E. Horvitz, "Managing human-robot engagement with forecasts and... um... hesitations," *Proceedings of the 16th International Conference on Multimodal Interaction*, 2014.

[121] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer, "Human-robot interaction in handing-over tasks," *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 107–112, 2008.

[122] F. Vannucci, A. Sciutti, H. Lehmann, G. Sandini, Y. Nagai, and F. Rea, "Cultural differences in speed adaptation in human-robot interaction tasks," *Paladyn, Journal of Behavioral Robotics*, vol. 10, pp. 256 – 266, 2019.

[123] K. R. MacArthur, K. Stowers, and P. A. Hancock, "Human-robot interaction: Proximity and speed—slowly back away from the robot!" in *Advances in Human Factors in Robots and Unmanned Systems*, P. Savage-Knepshield and J. Chen, Eds. Cham: Springer International Publishing, 2017, pp. 365–374.

[124] K. E. Schaefer, "Measuring trust in human robot interactions: Development of the "trust perception scale-hri"," in *Robust Intelligence and Trust in Autonomous Systems*. Springer, 2016, pp. 191–218.

[125] R. W. Bing and J. Michael, "Overcoming the uncanny valley through shared stressful experience with a humanoid robot," *Journal of Emerging Investigators*, 06 2018.

[126] M. Mori and K. F. MacDorman, "The uncanny valley: The original essay by masahiro mori," in *IEEE Spectrum*, 2017.

[127] A. Deshmukh, B. Craenen, A. Vinciarelli, and M. Foster, "Shaping robot gestures to shape users' perception: The effect of amplitude and speed on godspeed ratings," *Proceedings of the 6th International Conference on Human-Agent Interaction*, 2018.

[128] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, pp. 71–81, 01 2008.

[129] S. Donaldson and E. Grant-Vallone, "Understanding self-report bias in organizational behavior research," *Journal of Business and Psychology*, vol. 17, pp. 245–260, 2002.

[130] van de Mortel and F. Thea, "Faking it: Social desirability response bias in self-report research," *Australian Journal of Advanced Nursing*, vol. 25, pp. 40–48, 2008.

[131] D. Dodou and J. de Winter, "Social desirability is the same in offline, online, and paper surveys: A meta-analysis," *Computers in Human Behavior*, vol. 36, pp. 487–495, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563214002143

[132] K. Quigley, S. Kanoski, W. Grill, L. Barrett, and M. Tsakiris, "Functions of interoception: From energy regulation to experience of the self," *Trends in Neurosciences*, vol. 44, pp. 29–38, 01 2021.

[133] I. R. Kleckner, J. Zhang, A. Touroutoglou, L. Chanes, C. Xia, W. K. Simmons, K. Quigley, B. Dickerson, and L. F. Barrett, "Evidence for a large-scale brain system supporting allostasis and interoception in humans," *Nature human behaviour*, vol. 1, 2017.

[134] A. Craig, "How do you feel? interoception: the sense of the physiological condition of the body," *Nature Reviews Neuroscience*, vol. 3, pp. 655–666, 2002.

[135] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, pp. 169–200, 1992.

[136] K. Hoemann, M. Devlin, and L. Barrett, “Comment: Emotions are abstract, conceptual categories that are learned by a predicting brain,” *Emotion Review*, vol. 12, p. 175407391989729, 02 2020.

[137] L. F. Barrett, “The theory of constructed emotion: an active inference account of interoception and categorization,” *Social Cognitive and Affective Neuroscience*, vol. 12, pp. 1 – 23, 2017.

[138] A. Mehrabian, *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies.* Oelgeschlager, Gunn & Hain Cambridge, MA, 1980, vol. 2.

[139] A. Momen and E. Wiese, “Noticing extroversion effects attention: How robot and participant personality affect gaze cueing,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 62, no. 1, pp. 1557–1561, 2018.

[140] B. G. W. Craenen, A. A. Deshmukh, M. E. Foster, and A. Vinciarelli, “Shaping gestures to shape personality: Big-five traits, godspeed scores and the similarity-attraction effect,” in *AAMAS*, 2018.

[141] J. Ondras, O. Celiktutan, P. Bremner, and H. Gunes, “Audio-driven robot upper-body motion synthesis.” *IEEE transactions on cybernetics*, 2020.

[142] *A Robot that Uses Arousal to Detect Learning Challenges and Seek Help*, ser. ALIFE 2021: The 2021 Conference on Artificial Life, vol. ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems, 07 2014.

[143] D. S. Syrdal, K. Dautenhahn, S. Woods, M. L. Walters, and K. L. Koay, “’doing the right thing wrong’ - personality and tolerance to uncomfortable robot approaches,” in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006, pp. 183–188.

[144] Q. Xu, J. Ng, Y. L. Cheong, O. Tan, J. B. Wong, T. C. Tay, and T. Park, “The role of social context in human-robot interaction,” in *2012 Southeast Asian Network of Ergonomics Societies Conference (SEANES)*, 2012, pp. 1–5.

[145] H. Eysenck and S. Eysenck, “The eysenck personality inventory,” *British Journal of Educational Studies*, vol. 14, p. 140, 1965.

[146] H. Eysenck, “The structure of human personality,” in ., 1953.

[147] A. Tapus and M. Mataric, “Socially assistive robots: The link between personality, empathy, physiological signals, and task performance.” in *AAAI Spring Symposium - Technical Report*, 01 2008, pp. 133–140.

[148] A. Tanevska, F. Rea, G. Sandini, L. Cañamero, and A. Sciutti, “A socially adaptable framework for human-robot interaction,” *Frontiers in Robotics and AI*, vol. 7, p. 121, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2020.00121

[149] D. J. Haraway, “Situated knowledges: The science question in feminism and the privilege of partial perspective,” *Feminist Studies*, vol. 14, pp. 205–224, 1988.

[150] D. S. Syrdal, “The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study,” *Proc. New Frontiers in HRI Symposium, Edinburgh, 2009*, pp. 109–115, 2009. [Online]. Available: https://ci.nii.ac.jp/naid/20000716587/en/

[151] S. Yamada, T. Kanda, and K. Tomita, “An escalating model of children’s robot abuse,” in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 191–199. [Online]. Available: https://doi.org/10.1145/3319502.3374833

[152] L. Batrinca, B. Lepri, N. Mana, and F. Pianesi, “Multimodal recognition of personality traits in human-computer collaborative tasks,” in *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ser. ICMI ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 39–46.

[153] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro, “Multimodal recognition of personality traits in social interactions,” in *Proceedings of ICMI 2008*, 01 2008, pp. 53–60.

[154] R. Cattell, “The description of personality: basic traits resolved into clusters.” *The Journal of Abnormal and Social Psychology*, vol. 38, pp. 476–506, 1943.

[155] H. J. Eysenck and S. B. G. Eysenck, “Manual of the eysenck personality questionnaire,” in ., 1975.

[156] ——, “The eysenck personality inventory,” in ., 1965.

[157] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications." *Journal of personality*, vol. 60 2, pp. 175–215, 1992.

[158] J. Rogers and H. Christensen, "A conditional random field model for place and object classification," in *Proceedings - IEEE International Conference on Robotics and Automation*, 05 2012, pp. 1766–1772.

[159] I. Torre, A. Latupeirissa, and C. McGinn, "How context shapes the appropriateness of a robot's voice," *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 215–222, 2020.

[160] E. Coupeté, V. Weistroffer, O. Hugues, F. Moutarde, S. Manitsaris, and P. Fuchs, "New challenges for human-robot collaboration in an industrial context: Acceptability and natural collaboration," in *RO-MAN 2016*, 2016.

[161] P. Aliasghari, M. Ghafurian, C. L. Nehaniv, and K. Dautenhahn, "Effects of gaze and arm motion kinesics on a humanoid's perceived confidence, eagerness to learn, and attention to the task in a teaching scenario," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 197–206. [Online]. Available: https://doi.org/10.1145/3434073.3444651

[162] A. Cosgun and H. I. Christensen, "Context-aware robot navigation using interactively built semantic maps," *Paladyn, Journal of Behavioral Robotics*, vol. 9, no. 1, pp. 254–276, 2018. [Online]. Available: https://doi.org/10.1515/pjbr-2018-0020

[163] K. Nakadai, H. Okuno, and H. Kitano, "Robot recognizes three simultaneous speech by active audition," *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, vol. 1, pp. 398–405 vol.1, 2003.

[164] Z. Zafar, S. H. Paplu, and K. Berns, "Automatic assessment of human personality traits: A step towards intelligent human-robot interaction," *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 1–9, 2018.

[165] F. R. L. . F. M. R. . V. Matellan, "Context awareness in shared human-robot environments: Benefits of environment acoustic recognition for user activity classification," *IET Conference Proceedings*, pp. 24 (6 .)–24 (6 .)(1), January 2017.

[166] Y. Xiao, H. Liang, J. Yuan, and D. Thalmann, *Body Movement Analysis and Recognition*. Springer International Publishing, 09 2016, pp. 31–53.

[167] H. Liu, Y. Wang, W. Ji, and L. Wang, "A context-aware safety system for human-robot collaboration," *Procedia Manufacturing*, vol. 17, pp. 238–245, 2018, 28th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2018), June 11-14, 2018, Columbus, OH, USAGlobal Integration of Intelligent Manufacturing and Smart Industry for Good of Humanity. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978918311582

[168] M. B. Matthew Pan, Espen Knop and G. Niemeyer, "Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019.

[169] P. Gomez Esteban, E. Bagheri, S. Elprama, C. Jewell, H.-L. Cao, A. De Beir, A. Jacobs, and B. Vanderborght, "Should i be introvert or extrovert? a pairwise robot comparison assessing the perception of personality-based social robot behaviors," *International Journal of Social Robotics*, 03 2021.

[170] A. D. Kaplan, T. Sanders, and P. Hancock, "The relationship between extroversion and the tendency to anthropomorphize robots: A bayesian analysis," *Frontiers in Robotics and AI*, vol. 5, 2019.

[171] C. Jung and H. Baynes, "Psychological types," *Journal of Nervous and Mental Disease*, vol. 60, pp. 221–224, 1921.