FROG - Fun Robotic Outdoor Guide

Deliverable: **D4.1 part d**

Identification, evaluation and design of guide robot personality and behaviours: Design Guidelines for Robot Personality

Consortium

UNIVERSITEIT VAN AMSTERDAM (UVA)

IDMIND - ENGENHARIA DE SISTEMAS LDA (IDM)

UNIVERSIDAD PABLO DE OLAVIDE (UPO)

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE (ICL)

UNIVERSITEIT TWENTE (UT)

Grant agreement no. 288235

Funding scheme STREP

DOCUMENT INFORMATION

Project	
Project acronym:	FROG
Project full title:	Fun Robotic Outdoor Guide
Grant agreement no.:	288235
Funding scheme:	STREP
Project start date:	1 October 2011
Project duration:	30 September 2014
Call topic:	ICT-2011.2.1 Cognitive Systems and Robotics (a), (d)
Project web-site:	www.frogrobot.eu
Document	
Deliverable number:	D4.1 part d
Deliverable title:	Identification, evaluation and design of guide robot personality and behaviours - Design Guidelines for Robot Personality
Due date of deliverable:	m37 - 31 October 2014
Actual submission date:	M30 – 30 March 2014
Editors:	UT
Authors:	UT
Reviewers:	
Participating beneficiaries:	All Partners
Work Package no.:	4
Work Package title:	Socially Adaptive Human Robot Interaction
Work Package leader:	UT
Work Package participants:	UvA, Imperial, UT
Estimated person-months for deliverable:	7
Dissemination level:	Public
Nature:	Report
Version:	1 of part d
Draft/Final:	1
No of pages (including cover):	56
Keywords:	Human robot interaction; tour guide robot; robot personality, product personality, appearance; (set of) behaviors, design approaches, behavioral sciences, industrial design, online study, in-the-wild study

D4.1 D: DESIGN GUIDELINES FOR ROBOT PERSONALITY

CONTENTS

1. SUMMARY

In social robotics, personality for robots is used to give a robot consistent and understandable behavior. To date, most of the robot personalities have been based on inventories taken from human psychology. However, these have limited use for robots because not all dimensions of human personality can be made visible through robot appearance and behavior. We assume that the FROG robot should have an extravert personality, just like human tour guides have. The robot should also show openness to interact with people. Instead of basing the FROG robot personality on human behavior, we decided to design behavior for the FROG robot starting from a desired product personality. We expected that people would like a robot personality that was designed specifically for the modalities of the robot better than a robot that showed behavior copied from human tour guides.

Two personality options (consistent sets of behavior) were applied to the robot and tested for the robot. One was “human-like” and one was “modality specific.” For the human-like personality, behaviors copied from human tour guides were applied to the robot. For the modality specific personality, behaviors that were specially designed for the modalities of the FROG robot were applied. In both cases the set of behaviors was designed to form a consistent personality.

The two different personalities were used in two studies. First, we made a small 3D model of the FROG robot to create two short movies to be able to show the different personalities to a large number of people and to find differences in their experience of FROG robot personality. Second, the FROG robot was used for an in-the-wild study in the Royal Alcázar, showing one personality one day and the other personality the other day to the visitors.

From the questionnaire we found that participants found the robot with human-like personality more mechanical and more serious than the robot with modality specific personality. We think this means that participants found the robot with human-like personality more serious, because it distracted them less as they recognized the behavior from human behavior. This would indicate the use of a human-like personality for the robot. However, participants rated the human-like personality for the robot as more mechanical, indicating that copied human behavior also does not fit the robot. Participants remembered more of the story of the robot with human-like personality, which indicates that they seemed less distracted by a robot with human-like personality. This leads to our conclusion that copied human-like behavior does not fit a (mechanical robot), but that the behavior of the robot should not differ too much from human-like behavior, as modality specific behavior can be overdone and distract people from the story.

From the observations in the Royal Alcázar we found that the robot with human-like personality seemed to catch the attention of the visitors more easily, probably because the behavior is already known and familiar, but people also loose sooner their attention in human-like condition. The robot in modality specific condition seem to be better able to keep the attention of the visitors for at least two stops. When people follow more stops, there was no difference between the two

conditions, these people apparently were interested in the story and just follow the robot to obtain the information (which was exactly the same in both conditions). Also, we found that people who do not like tour guides seemed to like the robot with modality specific behavior very well, because this behavior does not remind of a human tour guide. In the modality specific condition the robot gave instructions on the screen, but some of the instructions given were not chosen well, because by asking people who already stand close to come even closer, one person might occupy the robot and exclude others from the tour.

The main result of the studies on defining the FROG robot personality is that the appearance of the robot seem to have more influence on the perceived robot personality than the various prepared robot personalities, but people do recognize or react differently to the different robot personalities. Also we found that people who do not like human tour guides, liked to receive some information of the robot with modality specific personality. The robot with human-like personality was better able to catch the attention, while the robot with modality specific personality kept the visitor attention for a longer time. In the modality specific condition the robot gave instructions to the visitors on what it expected them to do. People understood and followed the instructions, but some of the instructions were not well chosen. In that case one visitor stood very close and excluded other visitors from the tour. Overall it seems that for the target group – small groups of people who do not like to join human guided tours – a robot with modality specific behavior would fit best their preferences.

2. Introduction

Robots that perform social tasks are called social robots (Fong, Nourbakhsh, and Dautenhahn 2003). These social robots can be found in many places in our modern society. For example, robots can perform support tasks in health care and rehabilitation (Tapus and Matarić 2006), they can interact with older persons in elderly care (Kidd, Taggart, and Turkle 2006), they can perform education tasks, interacting with children (Mutlu, Forlizzi, and Hodgins 2006) and they can serve as innovative and interactive guides in museums or cultural heritage sites (Burgard et al. 1999; Clodic et al. 2006; Graf and Barth 2002; Nourbakhsh, Kunz, and Willeke 2003; Thrun et al. 1999). These robots are increasingly entering the "human world" and have to interact with people in a natural way. Therefore, they need to show socially accepted behavior and act according to socially accepted interaction rules.

As already stated in part a of deliverable 4.1, assigning a personality to the robot will help to create robot behavior that is consistent over time, is understandable and enables people to predict and interpret the actions and reactions of the robot (Duffy 2003; Kwak, Kim, and Kim 2008; Walters et al. 2007; Woods et al. 2005). In D4.1a (state of the art), we defined the term personality as: "T*he pattern of collective character, behavioral, temperamental, emotional and mental traits of an individual that have consistency over time and situations*" (taken from A. Tapus and M. Mataric) (Tapus and Mataric 2008). Or, as Fong et al. describe; personality is a set of distinctive qualities that distinguish individuals (Fong et al. 2003).

The main function of the FROG robot is to give an interactive guided tour. This requires interaction with the people around the robot. Therefore the robot has to show behavior that is socially accepted and understandable for people. We will design a personality for the FROG robot, which will help people to understand and predict the actions of the robot over time. As well as, the robot personality should be fun, so the tour will be fun and entertaining to follow.

In this deliverable guidelines for the robot personality will be given. Chapter 3 gives an overview of how and which personality inventories are used in social robotics and what kind of personality inventory we will use to design the FROG personality. In chapter 4, the different sets of behavior that form a personality are described. In chapter 5, an online study performed with stop motion movies of the robot (with the two different personalities applied) presenting two artworks is described. Chapter 6 present the results of a study with different robot personalities for a tour guide and people's reactions to the robot performed in a real-life setting (The Royal Alcázar). An overview of the guidelines in chapter 5 and 6, will be given in chapter 7 and in chapter 8 we describe how these guidelines will influence the work of the other partners. In chapter 9 we discuss the findings and in chapter 10 we give conclusions and recommendations for future work.

3. A ROBOT WITH PERSONALITY?

When you ask a person to describe his or her personality, or that of someone else, that person will come up with a set of characteristics such as: quiet or cheerful, thinker or talkative, serious or funny, easy-going or reserved. Even though these characteristics can describe a personality, they do not define the personality. In his book on Attitudes, personality and behavior, Icek Ajzen (Ajzen 2005) describes how these concepts are related and how people's behavior can be seen as indications of underlying personality traits. Throughout this deliverable, we will built on this understanding of how people's (and robot's) personality is reflected in their behavior. To date, there are several ways in psychology and behavioral sciences to measure human personality in a structured way. The characteristics are presented in questionnaires, to be filled out by people, and combined in an overarching dimension or trait. The Big Five and MBTI are two inventories that are often used to measure and define people's personality.

Personality in robots is used to make their behavior consistent over time and the actions of the robot better predictable and understandable for the people who are interacting with the robot (Duffy 2003; Kwak et al. 2008; Walters et al. 2007; Woods et al. 2005). There are many fields in social robotics that use robot personalities for their robot in interaction with people (for example: (Oh, Kwak, and Kim 2010; Tapus and Mataric 2008; Tapus and Matarić 2006; Walters et al. 2007; Woods et al. 2005).

One way to define robot behavior is to copy the behavior of humans. Yamazaki et al. developed behaviors for a robot tour guide by observing human tour guide behavior very specifically and by copying this behavior to a robot (Kobayashi et al. 2010). However, for more machine-like or simplified robots with limited degrees of freedom, these exact copies of human-like behavior can be very distracting for people, even though people seem to like the human-like behavior better than the more machine-like behavior (Karreman et al. 2013).

PERSONALITY INVENTORY SCALES

BIG FIVE

The Big Five personality inventory, also called the five-factor model consists of five broad dimensions that describe different factors of human personality. The Big Five dimensions are openness, conscientiousness, extraversion, agreeableness, and neuroticism, all of these consist of smaller factors based on (Goldberg 1999) (see Table 1). Over time, there have been several researchers who have used different definitions for the dimensions. For example the Big Three of Eysenck (also called EPI, Eysinck Personality Inventory) is based on the Big Five (Digman 1990) and uses the dimensions extraversion, neuroticism, psychoticism. The Big Five and EPI personality inventories are often used in human robot interaction research, for example, Lee et al. found that people were able to read the personality trait introversion-extraversion in the doglike robot AIBO, and enjoyed interaction with a robot with a personality complementary to their own more (Lee et al. 2006). Tapus and Mataric also used the introversion-extraversion trait when developing their robot behavior and tested which robot personality fitted which human personality best. In a rehabilitation context, they found that participants had a preference for personality matching (Tapus and Mataric 2008). Woods et al. explored whether robot personality should fit the human personality in workspace context and found that people preferred a robot personality that was different from their own personality (Woods et al. 2005). Walters et al. prepared three types of robot appearance and behavior and let participants evaluate the personality of the robots, they found that participants overall liked a human-like robot better, while introvert participants and participants with lower emotional stability tended to prefer a machinelike robot more (Walters et al. 2007).

Table 1: Dimensions and factors of the BIG FIVE personality scale

Openness	Intelligent, imagination, ideas, abstract thought, flexibility
Conscientiousness	Competence, self-discipline, deliberation, hard-working, helpful
Extraversion	Sociable, Warmth, assertiveness, activity, outgoing, confidence
Agreeableness	Trust, friendliness, straightforwardness, compliance, pleasant
Neuroticism	Emotional stability, anxiety, depression, impulsiveness

MBTI

The Myers-Briggs Type Indicator is another personality inventory, based on four dimensions of opposite concepts, dichotomies. The Myers-Briggs Type Indicator was developed by Katharine Cook Briggs and Isabel Briggs Myers and is a very popular measurement method in consultancy and training. The dichotomies Extraversion-Introversion and Thinking-Feeling are especially observable in the behavior of people and can be applied to robots. Based on these two dichotomies, Kim et al. (Kwak et al. 2008), proposed four personality types to use in human robot interaction; E-T, E-F, I-T, I-F (see table 2). In this study, the authors successfully controlled speed, velocity and frequency of robot gestures to influence the robot personality (Kwak et al. 2008). Kwak et al. used the MBTI to explore people's preference for robot personality and they found that robot temperaments are an effective indicator for constructing a personality for an entertainment robot (Kwak and Kim 2005).

Table 2: the four dichotomies of the MBTI

Extraversion (E)	Introversion (I)
Sensing (S)	Intuition (N)
Thinking (T)	Feeling (F)
Judging (J)	Perceiving (P)

Other ways of using personality for robots

Next to using personality inventory scales, researchers have applied other human personality characteristics to robots. For instance, Gockley et al. found that people preferred and had a less confusing interaction with a moody robot as opposed to a cheerful robot, because the actions of a moody robot were more predictable (Gockley et al. 2006). Goetz et al. found that people preferred a serious robot for a serious task (physical exercises) and a fun robot for a fun task (create a jellybeans recipe task) (Goetz, Kiesler, and Powers 2003), when they were performing tasks with the robot.

The preceding indicates that people recognize personalities in robots as they do in people, and that the behavior of a robot influences how its personality is evaluated. Moreover, a robot's personality can be developed to ensure that its behavior is perceived as consistent over time and to facilitate predictable and understandable actions (Duffy 2003; Kwak et al. 2008; Walters et al. 2007; Woods et al. 2005).

Personality inventory scales in HRI

As stated, the Big Five inventory and the MBTI measures (and inventories that are based on one of these) are often used in robot personality research. However, one disadvantage of these measures for research in robot personality is that not all dimensions/characteristics are easy to apply or to make recognizable in robots. For example, the dimension extraversion is used a lot, because it is found to be important for interpersonal interaction and therefore for Human-Robot Interaction as well (Lee et al. 2006). Also, agreeableness and openness are used in robot personality research and recognized by participants (Walters et al. 2007). However, of the Big Five dimensions, conscientiousness, neuroticism, openness and agreeableness are less easy to recognize compared to extraversion (Lippa and Dietz 2000). Therefore these personality characteristics are difficult to design for robots.

These inventories are often used in robot personality research to test which robot personality best combines with which human personality. The experiments done for this kind of research have mainly been performed with one individual person that had an interaction with one robot with a designed personality in a controlled setting (for example in (Tapus and Matarić 2006)). The measures lent from human psychology are very valuable for this kind of research and have been successfully applied to robotics. However, the situation that the FROG robot will function in is a different one. The FROG robot will be used by thousands of people over time and the robot will only interact with the visitors for a short period. Also, the FROG robot will not be used by one single person at the time, but by a group of people (who even might be strangers to each other) who might all have different personalities. In this case it will be hard for the robot to adjust its personality to the people it interacts with. In short, the personality for the FROG robot should be judged useful and helpful in the interaction by a lot of people and cannot be adjusted to one person's personality preferences.

We should add here that an interactive tour guide robot has to be an attraction for the visitors and should be the center of the attention for a while. The FROG robot has to actively attract the visitors and keep their attention for as long as possible, meanwhile making sure people can hear the robot and inviting people to interact with. These actions ask for active robot behavior, which fits with an extravert personality, just as the observed human tour guides (see D4.1b) had. Also, the robot should be perceived as intelligent, because people like to learn about the site they are visiting and will not follow a tour guide that is not giving serious information.

In the above explanation, two characteristics (extraversion and openness) taken from human psychology, are already defined for the FROG robot. This affects the usability of the Big Five or MBTI inventories, as they do not seem to be very suitable to use to define the FROG personality. Alternatively, we could use the concept of product personality to design a robot personality, because the robot is a designed entity. Product personality is used in Industrial Design to communicate about product appearance and functional characteristics, because designers use characteristics when designing products and consumers attribute characteristics when using products.

Product personality

As well as research on human personality characteristics, researchers have studied product personality or characteristics. A set of these characteristics together describe the product's perceived personality. As found by Mugge et al. a product can be described by 20 distinguishable dimensions (Mugge, Govers, and Schoormans 2009). These dimensions can be used by designers to design a personality for a product, but these dimensions can also be used to test whether the intended personality is recognized by the users. The 20 characteristics were found in iterative rounds of finding descriptive product characteristics and minimizing the amount of redundant characteristics.

The final characteristics that were found to identify a product personality are:

Cheerful	Cute	Obtrusive	Boring
Open	Idiosyncratic	Dominant	Aloof
Relaxed	Provocative	Untidy	Serious
Pretty	Interesting	Childish	Honest
Easy-going	Lively	Silly	Modest

The product personality scale was developed to identify the personality of products based on pictures or on real products. However, these can also be used to evaluate the experience of the interaction with the product (Desmet, Ortíz Nicolás, and Schoormans 2008). As people are also able to rate perceived experience with the product personality scale, we will use the scale to identify the experienced behavior of the robot, because behavior affects the perceived personality of robots.

When developing personality for the FROG robot, it might be more appropriate to use the items from the product personality scale to define the intended robot personality, than using the personality inventories of human psychology. We have determined that the robot should be fun, but serious, interesting, open, dominant and easy-going without being childish, boring, silly and untidy. These characteristics should be used when designing the appearance of the robot. Based on research by Schifferstein et al. (Ludden, Schifferstein, and Hekkert 2006) in which they show how the effect of visual, auditory, tactile and olfactory incongruence can lead to surprise or misunderstanding, we assume we can strengthen the characteristics visible in the design of the robot by designing the behavior and the interaction of the robot accordingly.

Research questions

Our main research question will be, how machine-like can a robot personality be and yet still be accepted in social tasks? Furthermore, what kind of behavior would fit this robot's personality best? With the research we describe here, we would like to find whether people prefer a simple robot with limited interaction modalities with a human-like personality, or, a modality specific personality. Therefore, we designed two types of behavior for our robot. In this research, we used different sets of behavior to influence the robot personality, as personality for a robot becomes visible through behavior and appearance. However, for this research we kept the robot design the same in both conditions, to find the effect of behavior on personality. For the behavior of the robot in the first condition, we studied human tour guide behavior and applied the observed behaviors to one robot. For the robot in the other condition, we designed a modality specific personality, meaning that the robot had a set of behaviors especially designed for the modalities of the FROG robot. This research design, performed online with videos of the robot and performed in the wild with the robot giving tours in the Royal Alcázar, will help us to answer the following questions.

Can and how does the behavior of the robot effect the personality that people attribute to the robot?

Can and how does the behavior of the robot effect how human-like or machine-like people think the robot is?

Do people prefer either a human-like personality or a modality specific personality for a tour guide robot, and, why?

Are people better able to pay attention to the art works when guided by a robot with modality specific behavior, and what causes the effect?

4. FROG ROBOT PERSONALITY

In robot development, the robot behaviors are often copied from human behavior, because people tend to understand these behaviors very well. And, as behavior influences the perceived robot personality, the personality will be understood as well. This seems to be true for robots that resemble humans a lot in appearance and modalities to communicate. For example, Yamazaki et al. found that copying gaze behavior of a human tour guide to a humanoid helped people to understand the story told by the robot better than when the robot showed random gaze behavior (Yamazaki et al. 2008). However we would like to argue that not all robots resemble people that much and for robots with small resemblance (and probably limited modalities to copy the human behavior) it might be more useful to use less human-like behavior. Behavior that is developed especially for the modalities the robot has, might be more easily understandable and less distracting for the people interacting with the robot.

To answer our research question, 2 studies were set out to test whether the different personalities -in this case only effected by the sets of behavior to test effects on the robot personality- influence people's understanding of the robot, the consistency of the robot personality and the user preference for the robot personality. To define the different personalities for the robot, a set of behaviors copied from human behavior was taken and applied to the robot and a set of behaviors specially designed to be optimal for the modalities of the FROG robot was created and applied to the robot.

As explained in the previous chapter, we did not change the robot personality on the scale of human psychology inventories, because the robot should be extravert and open because of its functionality. We did design nuances of behavior in each personality dimension. The behaviors for the robot in the two conditions were 1) based on the observed human tour guide behaviors and 2) based on the modalities that we were able to use of the robot. In this case the two conditions would keep the same level of extraversion and openness, while there were subtle differences in robot personality that were visible through behavior only.

DESCRIPTIONS OF ROBOT ACTIONS IN BOTH CONDITIONS

For the two studies the same sets of behavior (i.e. the human-like and the modality specific) that define the different personalities were used. The different sets of behaviors are presented in this section.

SET OF BEHAVIORS FOR HUMAN-LIKE PERSONALITY

The set of behaviors for the human-like personality was based on the observed behaviors of the human tour guides, as presented in Deliverable 4.1b. Not all of the behaviors as observed by human tour guides can be applied to the robot because the robot does not have the right modalities. Therefore, we had to make decisions for choosing the best modality to perform a behavior, or to leave a behavior out. For example, the robot does not have arms, therefore the pointer on the top of the robot was used to perform actions for which a human tour guide would use his or her arm.

Earlier studies showed that visitors liked a robot with human-like gaze behavior more than a robot that did not show human like gaze behavior. However, they were more distracted by the human-like behavior (see the study on gaze behavior in D4.1c and (Karreman et al. 2013)). We decided to use the observed human-like behavior set again (see Table 3, column 3) for this personality study, even though, we already knew people were distracted by human-like gaze behavior, because we do not know the outcomes for the other behaviors such as whole body movement (not for the purpose of pointing at an object) and pointing behavior by using a device (and not only the head). However, from (Karreman et al. 2013) we learned that the movements made to gaze at people or objects distracted people, therefore the "human-like gaze behavior" is minimized from the study described in (Karreman et al. 2013). The FROG robot cannot turn only its head, and the whole body movement is very distracting, so the robot does not turn too far to the objects.

Set of behaviors for modality specific personality

The set of behaviors for the modality specific personality was specially designed for the FROG robot. For these behaviors the interactional outcomes of the behavior of human tour guides were studied (for example, "steering the visitor attention to a specific part of the exhibit" is the interactional outcome of the behavior pointing). These interactional outcomes were desired outcomes of the robot behavior as well, using the modalities the robot is able to use. This led to a set of modality specific behaviors for the robot that is presented in Table 3 column 4.

Differences in the sets of behaviors

In Table 3 the different sets of behaviors for the human-like personality and the modality specific personality are described. To understand where these behaviors have their origins, the observed behavior of human tour guides as well as the interactional outcome are added to the Table. More details on human tour guide behavior and interactional outcomes can be found in deliverable 4.1b.

Table 3: Description of robot behavior in different personalities

Interactional outcome	**Human**	**Human-like personality** **Direct copied from human tour guide**	**Modality specific personality** **Optimal use of modalities to achieve interactional outcomes**
Catch the attention of the visitors	Make eye contact with the visitors Check busy and empty places Start with louder voice introducing themself	Turn on the spot as if looking at all visitors Start with introducing themself	Attention drawing eyes getting bigger, pulsing lights Pointer scanning for visitors, lights flashes on/off, follow visitors that come closer Screen says "come closer" "Press here for a guided tour"
Start at exhibit	Check if visitors follow and join Start with trivial information Pick one visitor to look at	Check if visitors gather around by turning on the spot Start with trivial information	Turn front towards place visitors should stand Start prerecorded voice (start

		Pick one visitor to look at	with trivial information) Eyes slightly pulsing Scan visitors with pointer and point to exhibit
Tell story *Keep visitors engaged and get implicit feedback* *Direct attention to the exhibit* *Direct attention to the exhibit*	Gaze at one person and change to the others sometimes Point in exhibit, supporting while saying "there" etc. Gaze to exhibit when pointing Show visuals from bag Depict story at exhibit Stand close to exhibit, but not in front of it (in front of pillar etc.) Orientation towards visitors	Keep gazing at one person, and by turning on the spot alternate to others Point into exhibit Gaze into exhibit while pointing Show visual on screen Show visual that is not visible in the site Stand close to exhibit but not in front of it Orientation towards the visitors	Show supporting terms on screen Scan visitors every once in a while with pointer Point into exhibit and eyes towards exhibit Show visual on screen Stand close to exhibit but not in front of it Orientation towards visitors
Indicate direction of next exhibit	Explain what is the next exhibit Point in direction	Explain what is the next exhibit Point in that direction	Explain what is next exhibit Point in direction with pointer
Finish at exhibit	Break eye-contact	Close eyes	Eyes start pulsing again Screen shows "please follow me" Pointer turns from visitors towards next direction
Start moving to next exhibit	Look in the direction of next exhibit Point towards next exhibit Start moving in direction of next exhibit	Turn slightly on the spot, as if looking at all visitors, then find small gap in between visitors to drive through Point towards next exhibit Start moving in next direction	Turn to next direction Keep pointing to next exhibit Screen shows "please follow me"
Guide visitors to next exhibit	Walk in front of group Talk to visitors in front Short distance	Drive in front of group Give some information about length of drive to front visitors	Drive in front of group Pointer keep pointing and sometimes scan for visitors, when looking at visitors light is pulsing Eyes are pulsing Screen shows "follow me"
Indicate having arrived at next exhibit	Stop waking Turn towards visitors again Look at all visitors	Stop driving Turn with front towards visitors again Turn on the spot to look at all visitors	Stop driving Turn towards visitors Eyes pulsing Pointer scan for visitors
End tour	Tell visitors tour is over	Tell visitors tour is over Eyes close	Tell visitor tour is over Eyes start pulsing less and less Pointer goes down to rest. Screen shows "this is the end of the tour"

5. Online evaluation of robot specific or human-like tour guide behavior

To evaluate different possibilities for the FROG robot personality, an online study was set up. The online study was used to collect a lot of quantitative data in a short time from participants all over the world. The results will be used to suggest guidelines for the FROG robot personality.

Based on our previous finding that people liked human-like robot behavior, but are distracted due to its movements (Karreman et al. 2013) and see D4.1c, we expected that people would rate the robot with the human-like personality as more anthropomorphic and like this robot better than the robot with a modality specific personality. However, we assumed that people would prefer the robot with the modality specific personality better because it would allow them to keep their attention on the artworks better and would not distract them with its movements, while it would still give the same amount of information and cues.

In the online study, participants were asked to rate the personality of a robot in one of two prepared movies. Since the visual aesthetics of the robot could not be manipulated for this project, we designed two alternative sets of behaviors for the robot that reflected different personalities, see chapter 4 for a detailed description of these two personalities. In the first condition, the robot showed a 'human-like' personality, while the robot showed a 'modality specific' personality in the other condition. During the online study after seeing one condition, people first answered a series of questions about the personality of the robot. Next, they were given a chance to see the other prepared robot personality (other condition). People were asked which personality they preferred and why. We found only subtle differences between the two personalities. A reason for this may be that people found it difficult to rate the different sets of behavior of a robot. The appearance of the robot might have a much larger influence on the perceived personality of the robot than the behavior does, similar to what was found by Walters et al. (Walters et al. 2007) when they compared a machine-like, neutral and human-like robot to be used as assistants in the house.

In this chapter, the method of the study will be described. The results of the study will be given as well as design guidelines that were drawn based on the results. Also, the results will be discussed on their relevance for the FROG robot. The results presented in this chapter and parts of the previous chapter are also used in (Karreman, Ludden, and Evers 2015a), which we submitted to CHI 2015.

Goal of the study

The goal of the online study was to test whether different types of behavior have an effect on how the personality of a robot is perceived and on how human-like people find the robot. Further, preference for a certain type of robot behavior was studied. As previously explained in chapter 2 we do not work with human personality traits for this study, because we assumed that a guiding robot should be extravert, open and intelligent. What we want to find out is how we should design the behavior of the robot so that the robot will be perceived as a fun robotic outdoor guide. To come to an effective fun robot tour guide, we assume the robot should be cheerful, open, interesting, serious and dominant, as these are characteristics that are important to be successful as human tour guides.

HYPOTHESES

This study was set up to give insight into how different types of robot behavior affect the perceived personality of a robot and into the considerations why people prefer one robot personality over another. The previous work and research questions lead to the following hypotheses:

H1: People will attribute a different personality to a robot with human-like behavior than to a robot with modality-specific behavior.

H2: People will find a robot with human-like behavior more human-like than a robot with modality-specific behavior.

H3: People will like the robot with the modality specific personality better.

H4: People are better able to keep their attention on the story when guided by the robot with modality specific behavior.

METHOD

An online study was set up in which people were asked to evaluate different types of robot behavior. For this study video-animations were used because the final robot was not available, yet. Using video-animations and an online questionnaire enabled us to collect the data during a short time span. Using video in HRI is an efficient manner to perform research, especially when people have to rate the personality or the behavior of the robot and they do not have a physical interaction with the robot (such as a hand-over task) (Woods et al. 2006). Using video is a suitable manner to collect data for this project, because it gives us the opportunity of getting feedback from a large number of people, which is not possible with lab or in-the-wild studies.

STIMULI

THE ROBOT

The robot used as tour guide in both movies is a 28-cm high model and simplified 3D model of the FROG robot. The robot was made of cardboard, hobby foam and colored paper. We made the robot look as much like the latest design of the FROG robot as possible. However, some simplifications were made to simplify the making of the animation. See Figure 1 for an impression of the robot design.

FIGURE 1: IMPRESSION OF THE ROBOT DESIGN

The following features of the robot were used to create different methods of communication: Animated eyes, pointer on top, visuals on screen, moving whole body and driving. To create the two different conditions (the human-like and the modality specific) these modalities were used in different ways.

The robot itself would not talk, but a voice-over was used for the explanations. The voice-over was a computer voice with speech generated by a text to speech engine. In the video, however, it might have seemed that the robot was "talking." The robot gave all the explanations in English.

Presentation of the stimuli

We prepared Two short movies (1:37 minutes and 1:39 minutes) showing the robot presenting two artworks in a museum setting. We designed two sets of behaviors that belonged to different personalities. In the first movie, the robot showed the designed set of behaviors for the human-like personality and in the second movie, the robot showed the designed set of behaviors for the modality specific personality. The exact descriptions of the different sets of behaviors that belong to the personalities can be found in Table 3 (chapter 4). Below, storyboards are given for both personalities to show the differences in the behavior sets.

Table 4: Storyboard of robot with human-like personality

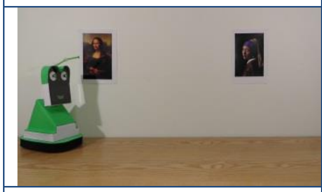
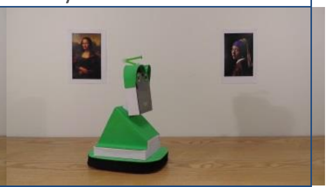
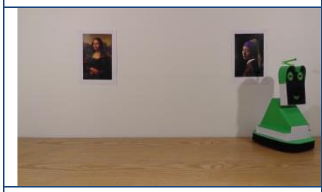
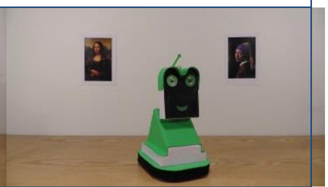
		
Start, welcome visitors and turn to others in semi-circle	Drive to first painting	Tell story of Mona Lisa, only subtle differences in the eyes
		
Point to Mona Lisa, turn body to painting	Give details, also turn to other visitors in semi-circle	Drive to next painting
		
Turn to visitors with orientation to environment	Tell story of the Girl with the pearl earring	End the tour

Table 5: Storyboard of robot modality specific personality

		
Start, information on screen and pointer checks for visitors	Drive, check for visitors behind it with pointer	Tell story of Mona Lisa, sometimes check for visitors with pointer
		
Point at Mona Lisa, information on screen and pointer to painting	Give details, with information on screen	Drive to next painting, keep pointing to end point.
		
Turn to visitors with orientation to paintings	Tell story of the Girl with the pearl earring	End the tour, giving information on the screen

At a first glance, differences between the sets may seem small. This is partly due to the fact that the story line is kept the same, but also due to the fact that the differences in behavior are subtle. The differences in behavior were in body movement, movement of the pointer, movement of the eyes and use of the screen. In these storyboards only stills are used but the behavior is dynamic. This makes it difficult to capture the dynamic differences in a few pictures.

The movies were controlled for amount of activity of the robot (see Table 5 for details). The amount of information about the artworks given in both conditions was the same; in both conditions the same sound files (with the robot explaining about the paintings) were used.

The paintings used in the story, the Mona Lisa and the Girl with the pearl earring, were chosen because these are quite well known paintings. However, the details given about the paintings were chosen so that they were most likely new to people, so the story would be interesting to listen to and so that we could test whether people had remembered the story from watching the video. The information was given in English. The voice used was a female computer synthesized voice.

Table 6: Activity of the robot in both conditions

	Turn	Drive	Search for visitors		Point to artwork		Show visuals	Change eye movement	
			Turn body	Use pointer	Turn to artwork and point	Use arrow on screen and point		Blink	Change animation
Human-like personality	4	4	7		7		1	12	
Modality specific personality	4	4		9		8	2		7

Participants

A total of 204 participants (132 males, 58 females, 14 preferred not to disclose, aged 15-58, mean 31,6) evaluated the stimuli. All participants were members of Crowdflower. The participants came from different continents all over the world (most from Europe 42% and Asia 31%).

Procedure

The videos and questionnaires were placed on Surveymonkey. A link to the videos and questionnaires was placed on Crowdflower, which is a crowdsourcing website. On this website everyone can register to fulfill tasks such as questionnaires. Crowdflower participants chose to do the work-unit and were redirected to one of two questionnaires on Surveymonkey. We used two different questionnaires, one showing the video with the robot with a human-like behavior first, the other showing the robot with a modality specific behavior first. At the start of the study, on page one of the questionnaire, the participants were asked for their consent. By giving consent, they agreed that the researchers would use the obtained data for academic research and academic and general publications.

The online questionnaire started with a video of either the robot with the human-like behavior or of the robot with the modality specific behavior on page two. The participants were randomly assigned to one of the two conditions.

After seeing the video, the participants were asked to fill in a questionnaire that asked about the perceived personality of the robot and about their evaluation of the behavior of the robot. All items in the questionnaire were presented to the participants on 5-point Likert-scales or 5-point semantic differentials, unless stated otherwise.

On page three, constructs of the Godspeed questionnaire and the Source Credibility Scale and four items on obviousness, novelty, being qualified and reliability of the robot were used. The constructs of the Godspeed used were Anthropomorphism, Likeability, Animacy and Perceived Intelligence. The used constructs of the Source Credibility Scale were Sociability, Extraversion, Competence and Character. Some items were asked once in the questionnaire, but used in two different constructs during analysis. This is the case for "artificial-lifelike" (used in Anthropomorphism and Likeability), "unfriendly-friendly" (used in Likeability and Socialbility), and "unintelligent-intelligent" (used in Perceived Intelligence and Competence). Four items were added to mask the intention of the questionnaire. See Table 6 for the items used in the various constructs.

TABLE 7: CONSTRUCTS OF GODSPEED, SCOURCE CREDEBILITY SCALE AND HOME MADE QUESTIONS

Godspeed – Anthropomorphism Fake-Natural Machinelike-Humanlike Unconscious-Conscious Artificial – Lifelike Moving rigidly-Moving elegantly	**Godspeed – Animacy** Dead – Alive Stagnant – Lively Artificial – Lifelike Inert – Interactive Apathetic – Responsive Mechanical - Organic	**Godspeed – Likeability** Dislike – Like Unfriendly – Friendly Unkind-Kind Unpleasant-Pleasant Awful – Nice
Godspeed – Perceived Intelligence Incompetent – Competent Ignorant – Knowledgeable Irresponsible – Responsible Unintelligent – Intelligent Foolish – Sensible	**Source Credibility Scale – Sociability** Irritable – Good-natured Gloomy - Cheerful Unfriendly – Friendly	**Source Credibility Scale – Extraversion** Timid – Bold Quiet – Verbal Silent - Talkative
Source Credibility Scale – Competence Inexpert – Expert Unintelligent – Intelligent Narrow - Intellectual	**Source Credibility Scale – Character** Dishonest – Honest Unsympathetic – Sympathetic Bad – Good	**Home made** Unclear - Obvious Ordinary-Novel Unqualified-Qualified Unreliable - Reliable

On page four of the questionnaire, the participants were asked to rate the behavior of the robot on a product personality scale. The product personality scale (adapted from (Mugge et al. 2009)) consisted of 20 items that were presented in a 5-point Likert scale with end points strongly disagree and strongly agree, see Table 8 for all items.

TABLE 8: PROCUCT PERSONALITY SCALE

Cheerful Open Relaxed Pretty Easy-going	Cute Idiosyncratic Provocative Interesting Lively	Dominant Obtrusive Childish Untidy silly	Boring Aloof Serious Honest modest

After this, on page five, 9 questions asking whether participants had understood the behavior of the robot were given. On page six, the participants had to answer 4 multiple choice questions (4 options, of which one was always "I can't remember") tested whether people had remembered the story that the robot told (see Table 9).

TABLE 9: UNDERSTANDING OF THE ROBOT

It was easy for me to understand the story of the robot
I knew what the robot talked about
The story of the robot interested me
The story the robot told was clear to me
I knew when I was supposed to look at the painting
I knew where I had to look in the painting
It was easy for me to keep my attention at the paintings
The robot distracted me from the paintings
The robot often got my full attention

At this point, the participants were given the option to see the video that presented the other condition. After seeing this movie, the participants were asked on page nine, which robot behavior they liked best (2 option multiple choice question), and they were asked 3 open questions about why they prefered a specific behavior, what the main difference between the robot behavior in the two movies was and whether they had any suggestions for the robot behavior.

All participants were asked for personal details (page ten) such as age, gender, education level, work field, experience with social robots and the like. When they had finished the questionnaire, the Crowdflower members received a code to obtain a small payment for their participation.

DATA ANALYSIS

After preparing the data, a total of 204 participants were left. The participants were evenly divided over the two conditions, however due to the fact that some participants were left out of the analysis, we had 114 cases for the robot with human-like personality, and 90 cases for the robot with modality specific personality left to analyze. Participants were left out when they stopped after the first question or halfway through the questionnaire; after that the collected answers to the questionnaires were carefully examined on honest participation, the people that gave the same answers everywhere were taken out; and the participants that filled out the questionnaire in less than 7 minutes were left out the analysis (a total of 29 participants), as this is too short to finish the questionnaire. Of all participants, 167 decided to take part in the non-mandatory part of the questionnaire and observed both robot behavior conditions.

The items of the several constructs and other questions in the questionnaire were randomized per page. The questions on page nine and ten were not randomized, as their order was important for answering the questions.

The used scales were checked for reliability, using the Cronbach's alpha. ANOVA's were used to test for differences in perceived personality of the robot between the different conditions. Also, ANOVA's were used to test interaction effects between condition and other factors for differences in perceived personality.

We used a factor analysis to check whether the self made items of the "attention" scale contributed to one or more factors. 7 of the 9 items contributed to one factor, of which a construct was made and used for further analysis.

Results

Reliability of measures

In this section, the Cronbach's Alpha we found for the several constructs are presented. A Cronbach's Alpha above 0.7 is considered as reliable. However, as we used small scales, the Cronbach's Alphas we found tend to be a little low.

Table 10: Cronbach's Alpha for constructs

Measurement	Number of items	Found Cronbach's Alpha coefficient
Godspeed – Anthropomorphism Fake-Natural Machinelike-Humanlike Unconscious-Conscious Artificial – Lifelike Moving rigidly-Moving elegantly	5	*0.693*
Godspeed – Animacy Dead – Alive Stagnant – Lively Artificial – Lifelike Inert – Interactive Apathetic – Responsive Mechanical - Organic was left out to increase the Cronbach's Alpha.	5	*0.694*
Godspeed – Likeability Dislike – Like Unfriendly – Friendly Unkind-Kind Unpleasant-Pleasant Awful – Nice	5	*0.825*
Godspeed – Perceived Intelligence Incompetent – Competent Ignorant – Knowledgeable Irresponsible – Responsible Unintelligent – Intelligent Foolish – Sensible	5	*0.695*
Source Credibility Scale – Sociability Irritable – Good-natured Gloomy - Cheerful Unfriendly – Friendly	3	*0.724*
Source Credibility Scale – Extraversion	3	*0.444*

Timid – Bold Quiet – Verbal Silent - Talkative		
Source Credibility Scale – Competence Inexpert – Expert Unintelligent – Intelligent Narrow - Intellectual	3	*0.672*
Source Credibility Scale – Character Dishonest – Honest Unsympathetic – Sympathetic Bad – Good	3	*0.674*
Understandability of robot behavior It was easy for me to understand the story of the robot I knew what the robot talked about The story of the robot interested me The story the robot told was clear to me I knew when I was supposed to look at the painting I knew where I had to look in the painting It was easy for me to keep my attention at the paintings	7	*0.845*

RESULTS

H1: ATTRIBUTE DIFFERENT PERSONALITY TO THE ROBOTS

Our first hypothesis stated that people will attribute a different personality to a robot with human-like behavior than to a robot with modality-specific behavior. This hypothesis also checks whether the manipulation of the behavior was done the right way and whether people found a difference between the two designed robot personalities (did the participants of both conditions rate the different sets of behavior differently?). We assumed that participants would rate the robot with the human-like personality higher on anthropomorphism than the robot with the modality specific personality.

To test this hypothesis, we performed an independent sample T-Test with condition as independent variable. The dependent variables were the Godspeed and Source Credibility scales, all individual items of the Godspeed and the Source Credibility scales, as well as the individual items of the Product Personality scale.

In our T-Tests, we found no significant difference between the participants' scores for the two robot personalities on the Godspeed or Source Credibility Scale constructs. However, there was a difference on the item Mechanical - Organic from the Godspeed scale *t(202)=6.341, (p=0.000).* Participants rated the robot with the human-like behavior as more mechanical, the robot with the modality specific behavior was rated as significantly more organic. Also, the robot with the human-like behavior scored significantly higher on 'serious' from the product personality scale than the robot with the modality specific behavior *t(202)=2.269, (p=0.024).* These results indicate that there may be very subtle differences in how people relate to human-like and modality specific designs. Therefore, we conclude that our first hypothesis is only weakly supported.

The results presented above might suggest that our manipulation was not successful. However, after seeing both videos, 95 participants (57.9% of the participants who saw both videos) stated they saw differences in the robot behavior. 34 participants (20.7%) stated that they did not see differences, and 35 (21.3%) gave an answer unrelated to the question. The condition, age or nationality was not found to influence these results.

H2: People will find a robot with human-like behavior more human-like than a robot with modality-specific behavior.

We assumed that people would find the robot with the human-like personality more human-like than the robot with modality specific personality, because the set of behaviors was copied from human tour guides and people would recognize the behavior. To test this hypothesis, we performed an independent sample T-Test with condition as independent variable; the construct Anthropomorphism of the Godspeed questionnaire and the individual item human-likeness were dependent variables. We did not find any significant results on human-likeness and anthropomorphism. Therefore, we cannot support this hypothesis.

H3: People will like the robot with the modality specific personality better.

Hypothesis 3 concerned the likeability of the robot. We assumed that people would like the robot with modality specific personality better, because the robot with this personality is clearer in where to look and moves less so less distractive. To test this hypothesis, we performed an one-sample T-Test with the "preferred condition" as variable to test if one of the conditions was preferred, but we did not find a significant difference. We then performed an independent sample T-Test with condition as independent variable and "preferred condition" as dependent variable to test if the order of viewing the different videos might have influenced participants' preferences. This test did not yield a significant result either. Additionally, we compared the number of people that preferred one condition over the other after they had seen both videos. Of the participants that saw both types of behavior, 76 participants (45.8%) preferred the robot with the human-like behavior, while 90 participants (54.2%) preferred the robot with modality specific behavior. The preference for one of both robot behaviors might be a matter of personal opinion, this was also reflected in the answers people gave to the open questions. The order of viewing the different videos with different robot personality did not influence participants' preferences for a set of robot behavior. Based on these results we conclude that this hypothesis was not supported.

H4: keep their attention on the story

The fourth hypothesis concerned the attention that people had for the story that the robot told. We expected that people would be better able to keep their attention on the robot with the modality specific personality.

To test this hypothesis, we performed an independent sample T-Test with the items of the attention scale as dependent variable and the condition as independent variable. There were no significant differences in the self-report of how participants understood the robot's stories or participants' attention for the artworks.

However, we also analyzed to what extent people remembered the information provided by the robot by asking some detailed questions about the content. For this analysis, we counted correct answers per participant, incorrect answers per participant, and "I can't remember" answers per participant. We then used independent sample T-Tests to check for differences between conditions. We found that people who saw the human-like condition gave significantly more correct answers than people who saw the robot in the modality specific condition $t(201)=2.018$, ($p=0.045$). And vice versa, participants who saw the robot with the modality specific behavior gave significantly more incorrect answers than participants who saw the robot with human-like behavior $t(201)=2.562$, ($p=0.011$).

These results do not support hypothesis 4, but rather indicate the opposite: it seems that participants who saw the robot with the human-like behavior were better able to keep their attention to the story the robot told and gave significantly more correct answers. Answers to open questions did not give any indication that participants might have been distracted by one of the sets of behaviors.

Findings not related to the hypotheses:

During the analysis, we found some things that were not related to the hypotheses, but that do influence the visitor experience of following a robot tour guide.

First, in the open questions, participants often remarked that the voice of the robot was too mechanical and not nice to listen to. The participants suggested changing the voice to a more human voice, which is easier to understand.

Second, the scores the participants gave on the Product Personality scale were very close to each other for both prepared robot personalities, as can be seen in Figure 2. A significant difference was found only for the scores on the item 'serious,' which is explained in the previous section. The fact that the scores on the items on the Product Personality Scale are similar to each other means that we cannot conclude that the robot with humanlike personality was perceived significantly different from the robot with the modality specific personality.

As the results of the rating of both robot behaviors are very close to each other, we looked at the aggregated responses on the scales that asked about the personality of the robot to test how people perceived the robot personality independent of the behavior. From this analysis we found that the robot product personality was perceived as cheerful, relaxed, easy-going, cute, lively and modest (all had a score above 3.4). Also, the robot personality was perceived to be interesting and honest, with a score above 3.7. The robot was not perceived as provocative, obtrusive, untidy, silly and boring (all scores below 2.8).

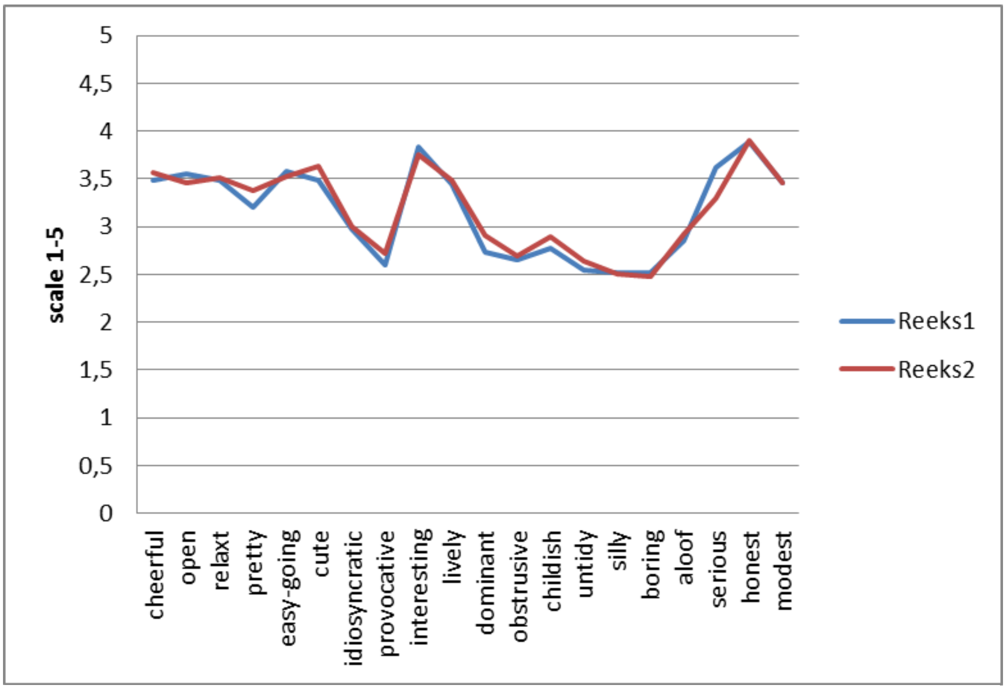


FIGURE 2: FROG ROBOT PRODUCT PERSONALITY

DISCUSSION

This study was set up to find differences in participant's experience of and preference for a robot personality. However, from our manipulations of robot behavior, only subtle differences in perceived personality were found. In this section we will discuss the results and the possible explanations why these results were found. Also, some implications of the design of the experiment will be discussed.

The robot with the human-like personality was rated as more mechanical and the robot with the modality specific personality was rated as more organic (*t(202)=6.341, (p=0.000))*. This is probably because the human-like behavior was copied to the robot, but as not all behavior observed from the tour guides could be copied to the robot, it might not have been as natural for the robot as it is for human tour guides, and therefore is perceived as more mechanical. The modality specific behavior, on the other hand, was developed especially for the robot, and therefore it might be perceived as more smooth and organic behavior for the robot.

The robot with human-like personality was perceived as more mechanical and the robot with the modality specific behavior was perceived as more organic, therefore, we argue that the modality specific behavior fits the FROG robot better and the human-like behavior does not fit the robot. This indicates that using a modality specific personality for the FROG robot would be advisable.

On the product personality scale, we found less difference between the both manipulated behaviors for the different characteristics. However, there is a significant difference for the item 'serious'. The robot with the human-like behavior scored significantly higher on this item than the robot with the modality specific personality (*t(202)=2.269, (p=0.024))*. This might be because people find the behavior of the robot with human-like behavior easier to understand, because they relate it to their experience with human tour guides. Also, the robot with modality specific behavior seemed to distract people more, as people who saw the robot with modality specific behavior gave more wrong answers to the detailed questions about the robot's story. This might be because they did not have previous experience with that robot behavior, and had to pay attention to the behavior of the robot to understand its actions.

This finding will have influence on the design of robot behavior. When designing robot behavior in modality specific style, the behavior should not deviate too much from what people are familiar with. The modality specific personality of the robot was probably perceived as "strange" or "unknown," which people might perceive as distracting and therefore "less serious." From this we learned that a robot personality should not be too unfamiliar, while a direct copy of human-like personality is also not preferable.

Here we should note that the movies only took about 1.40 minutes. Therefore, people may not have had enough time to learn to understand certain behavior. This might explain why people liked the robot with human-like personality better, because people are more familiar with this behavior. However, when people have the opportunity to spend more time with the robot with a modality specific personality, the preference for robot personality might be different. This needs to be tested in more extended follow-up studies.

There might be several reasons why only very subtle differences between the robot with the human-like personality and the robot with the modality specific personality were found. Below we give some reasons that might explain this.

First, we will discuss the main reason why we found only subtle differences. Participants in both conditions rated the robot behavior more or less the same, even when people answered in the open questions after seeing the other movie some or more differences in the behavior. Most people had a preference for one of both movies, based on the robot behavior sets (and thus the personality). This means people see the differences, but only when they have seen the two movies. People seem to rate the design and behavior as one, as they cannot really separate these two. The fact that the scores for the robot product personality for both prepared personalities are so close to one other supports this. This could indicate that design of the robot has a higher influence on the perceived personality of the robot than the sets of behaviors. Alternatively, it could be argued that, although we tried to design two types of behaviors that were different, we also wanted the general experience of the robot as an open and honest guide to be maintained. Therefore, the eventual manipulations of behavior were probably too similar to lead to differences in perceived personality. From this we conclude that it is very important to design the robot personality (set of behaviors) in such a way that it fits with and strengthens the robot appearance.

Second, the participants were asked to rate the personality of the robot after watching a video. However, a personality best becomes clear in interaction (through behavior). Participants in this study were not interacting with the robot. As a result, they might not have experienced the personality of the robot as intensely as they would have while interacting with the robot. This might have influenced ratings on the personality profiles of the robot.

Third, in this online study 204 people from all over the world participated (41.7% from Europe, 30.9% from Asia, 11.3% from North America, 5.9% from South America, 2.9% from Africa and 0.5% did not answer the question). All these people may have different opinions towards robotics and convenient behaviors for robots. We indeed found significant differences for perceived personality of the robot between people from different counties, however, this was on items only and not on full scales.

We want to discuss two limitations in our online questionnaire study here. First, the 204 participants that participated in the study were not checked for their own personality profile, which might influence how people perceived the interaction. We did not test for participant personality because when the robot guides people in the Royal Alcázar, it will not change its behavior set based on the personality of the people it is interacting with either. Therefore, we assumed that we have to find a robot personality that attracts all people regardless of their own personality profile. Second, as also stated earlier in this chapter, the two storyboards of the robot personalities that were presented look very similar. Only subtle differences in the personality were made. This could also influence the fact that there are small differences in the results.

To conclude the discussion, we can state that we were able to design different robot personalities by using different sets of behaviors. However, the differences found are rather small, which was also expected, as we did not have differences in the appearance of the robot and nor did we have differences in the personality dimensions extraversion or intelligence. We found that the appearance influences the personality of the robot to a large extent and that subtle differences in behavior only results in subtle changes. We found that the robot with the modality specific personality was perceived as being more organic (than mechanical) and less serious than the robot with the human-like personality.

Guideline for robot personality

In the study described we asked people to evaluate the personality of one of two guide robots with different personalities; a human-like personality and a modality specific personality. We found that a human-like personality for a robot was perceived as mechanical and did not fit the tour guide robot. This finding is in line with our previous study on gaze behavior for a museum robot, where we applied human-like gaze behavior to a robot, which people did not like either (Karreman et al. 2013). Now we found that people also did not like copies of human-like behaviors for the other modalities of the robot, which strengthens our hypothesis that a modality specific personality is more effective for a tour guide robot.

The modality specific behavior was designed especially for the modalities of the robot, and better fitted the function of the robot. However, a modality specific personality for a robot was perceived as unfamiliar, which made the robot distracting and therefore less serious. We assume this will be solved if people have longer time to get used to the behavior. Still, a personality that is unfamiliar from the known, distracts people from the content, which is not the intention of a tour guide robot. Therefore, the modality specific robot personality should not deviate too much from known human-like behavior, or at least the interactional outcomes of the tour guides should be aimed for when designing the behaviors, then, we assume, the intentions of the robot will be recognizable.

Based on the results found in the online study, we assume that the appearance of the robot has a larger influence on the robot personality than the different designed personalities do. This is in line with the findings of Walters et al. (Walters et al. 2007). Therefore, we advise to define a personality profile for the robot before making the visual appearance and then to make the appearance of the robot that is according to the profile. The behavior can then be designed according to the same personality profile to make sure that the behavior and appearance strengthen each other.

While people liked the modality specific personality best for a robot, the voice of the robot should not be mechanical or machine-like as the other modalities. People prefer a human voice over a mechanical voice, because a mechanical voice is difficult to follow and understand and also slow at times. The robot voice should preferably be a human voice, which will improve the understandability.

6. Real world evaluation of robot specific or human-like tour guide behavior

On top of the evaluation of different options for the FROG robot personality in the online study, an in-the-wild study with naïve visitors was set up. This study was used to collect qualitative data of reactions of people in interaction with the FROG robot that showed one of two different personalities, which were the same as in the online study. The results of this study will be compared to and combined with the results of the online robot personality study to develop guidelines for the FROG robot personality.

The real-world study was performed in the Hall of Festivities in the Royal Alcázar in Seville during the FROG integration week in May 2014. During the study, the robot gave short tours with five stops of approximately 20 seconds each. Three sessions of short guided tours were performed, one with the robot showing human-like behavior and two with the robot showing modality specific behavior. See chapter 4 for a detailed description of the two personalities that form the basis for these behaviors. Naïve visitors could join a tour of the robot if they liked, but they were not forced to. Also, they could follow the robot as long for as or as short as they preferred. Stills of sequences of the video recordings of these tours were used to analyze the visitor reactions to the two different robot behaviors. Also, during the sessions in the Royal Alcázar, 17 individuals or pairs were interviewed after they had joined the robot tour for one or more exhibits. They were asked for their experience of following the robot tour guide.

We found that people who do not like tour guides seemed to like the robot with modality specific behavior very well, probably because this behavior does not remind of a human tour guide. However, the robot with human-like personality seemed to catch the attention of the visitors more easily, probably because the behavior is already known and familiar. On the other hand, people in the human-like condition more often followed just one stop than people in the modality specific condition, while people in the modality specific condition more often than people in the human-like condition followed the robot for two stops. The modality specific personality for FROG therefore seems to be more effective, but some instructions, such as “come closer” should be given in a smarter way, as now one person by coming very close excluded others from the tour.

In this chapter, the method of the study will be described. For the analysis, the data of the moments that the robot explains something at a Point of Interest and the interview data of the 17 short interviews were used. The results of the study will be given as well as design guidelines that were drawn up based on the results.

Goal of the study

The goal of this “in-the-wild” study with the robot tour guide FROG in the Royal Alcázar was to check whether different types of behavior of the robot influenced the reactions and behaviors of visitors in real life compared to people’s experiences when rating robot behavior show in a video. Furthermore, preference for a certain type of robot behavior was studied by interviewing people who joined the robot tour guide. In this study the same two robot personalities as in the previously described online study were used and we wanted to find what kind of behavior would best fit the FROG robot for effectively guiding visitors through an indoor/outdoor cultural heritage site.

Hypotheses

This study was set up to give insight into how different types of robot behavior affect the reactions and behavior of people that encounter the guide robot. We expect that one of both sets of behavior will be more effective in guiding, for example in focusing the visitor's attention on a Point of Interest, the number of stops people follow and the formations they form around the robot.

Based on previous research we came to the following hypotheses:

H1: People will react differently to a robot with human-like behavior than to a robot with modality-specific behavior, for example in their way of listening to the robot and following the robot, their gestures to the robot, their formation around the robot, if they take pictures and pose with the robot, as well as in their answers to interview questions.

H2: People will prefer the robot with the modality specific personality as a robot tour guide, because it makes its intentions more clear.

H3: People will react to the robot with the modality specific personality as they would do to a human tour guide more than they will react to a robot with human-like behavior, because it makes its intentions more clear.

H4: People will be better able to keep their attention on the story when guided by the robot with modality specific behavior and will therefore follow the robot for more stops.

Method

The in-the-wild-study was set up in the Royal Alcázar in Seville, Spain. In three different time slots, the robot performed short guided tours with two different sets of behavior. Visitors that encountered and joined the robot tour guide were observed for their reactions, behavior and interactions with the robot and some of the visitors were randomly picked out to participate in a short interview.

Stimuli

The robot

The robot that was used for the study was the FROG platform, see figure 2, on which different sets of behavior were programmed to be able to show the two different personalities. The FROG robot was controlled remotely by one of the project members from IDMind. The controller of the robot was the same for all sessions performed during the study.

The robot was able to interact with visitors using its body movements, the eyes (sequences played with the LED-lights in the eyes), the pointer (pointing at things and sequences played with the LED-lights in the pointer), a touchscreen that was used to show pictures and movies, and prerecorded speech. In this experiment the visitors were not asked to use the touch screen, also, the robot would not understand the visitors if they would talk to it.

Figure 3: FROG in the Hall of Festivities

The robot used for the study in the Royal Alcázar had power to drive around for approximately one and a half hours. Afterwards the robot needed to recharge for several hours. Therefore, the study was performed in time slots of a maximum of one and a half hours each. During the study, the robot tried to catch the attention of visitors and gave several tours of approximately three minutes. One researcher interviewed some of the participants briefly after they had followed a (part of a) tour.

Presentation of the stimuli

One short tour was prepared for the study, while two different robot personalities were prepared. In the first condition the robot communicated its intentions with human-like behavior, in which human-like behavior was copied as well as possible to the robot (e.g. the robot turned towards the exhibit, looked towards the exhibit, pointed in the exhibit). In the other condition the robot had a modality-specific behavior, in which the behavior was adjusted to the modalities the robot had to communicate its intentions (e.g. the robot used the pointer to check engagement of visitors, the screen was used to attract attention, the robot did not turn towards the exhibit). In both conditions, the robot presented the same exhibits and the same text, but the behavior to support the text was different. The exact descriptions of the different sets of behaviors that belong to the personalities can be found in Table 3 (chapter 4).

Below, storyboards are given for both personalities to show the differences in the behavior sets. For the same reasons as in the online study, the differences between the sets may seem small. This is due to the fact that the story line was kept the same and the differences were mainly visible in the body movement, movement of the pointer, movement of the eyes and use of the screen, which makes it difficult to capture the dynamic differences in several pictures. In Table 1 and Table 2 are the different sequences and descriptions of the robot behavior given for both conditions. The sequences per exhibit are taken from different tours, to give a good overview. The differences between the two exhibits are not visible in the position of the robot, although these differed a bit between the sessions due to the people that were in the room. From these pictures it is probably difficult to understand what a tour looked like, because the differences in behavior of the robot are very small and mainly visible in the screen and pointer actions.

TABLE 11: STORYBOARD OF ROBOT SHOWING HUMAN-LIKE BEHAVIOR

		
Welcome – The robot turned its whole body to visitors that walked by. On the screen a mouth was visible.		
		
Explanation about the design of the tiles – The robot turned its whole body a bit to the different visitors while giving information. On the screen a mouth was visible and when the robot pointed at the exhibit, the pointer pointed there and the robot turned its whole body slightly in that direction.		
		
Explanation about banner above the door - The robot turned its whole body a bit to the different visitors while giving information. On the screen a mouth was visible and one explaining picture was shown for a few seconds. When the robot pointed at the exhibit, the pointer pointed there and the robot turned its whole body slightly in that direction.		
		
Explanation of the faces of Charles and Isabel - The robot turned its whole body a bit to the different visitors while giving information. On the screen a mouth was visible and one explaining picture was shown for a few seconds. When the robot pointed at the exhibit, the pointer pointed there and the robot turned its whole body slightly in that direction.		
		
Goodbye - The robot turned its whole body to visitors that stood close. On the screen a mouth was visible.		

Table 12: Storyboard of robot showing Modality specific behavior


Welcome - The robot stood in a position and used the pointer to search for people. On the screen the text “welcome!” and “come closer” was visible, during the explanation the screen was black, and at the end of the explanation the screen showed “Follow me!”

Explanation about the design of the tiles - The robot stood in a position and used the pointer to search for people. Before the explanation started, the screen showed “come closer.” During the explanation the screen was black, except when explaining where to look by using a picture and markers on the screen. The pointer was used to point at the exhibit. After the explanation, the screen showed “follow me.”

Explanation about banner above the door - The robot stood in a position and used the pointer to search for people. Before the explanation started, the screen showed “come closer.” During the explanation, the screen showed an arrow to indicate where to look, also an explaining picture was shown for a few seconds. When the robot pointed at the exhibit, the pointer pointed in that direction. After the explanation, the screen showed “follow me.”

Explanation of the faces of Charles and Isabel - The robot stood in a position and used the pointer to search for people. Before the explanation started, the screen showed “come closer.” During the explanation the screen was mainly black, only one explaining picture was shown for a few seconds. When the robot pointed at the exhibit, the pointer pointed in that direction. After the explanation, the screen showed “follow me.”

Goodbye - The robot stood in a position and used the pointer to search for people. During the explanation the screen was black, and in the end of the explanation the screen showed “Goodbye!”

The two behaviors were controlled for amount of activity of the robot. While the robot with human-like behavior was more dynamic in its body movement, the robot with modality-specific behavior gave more information on the screen and used the pointer more. Due to the real-life setting a specification of the exact balance of robot behavior is difficult to give as it also was influences by the reactions of the visitors, for example, when visitors stood in front of the robot, the robot needed to find a way to pass in both conditions.

PARTICIPANTS

All visitors of the Royal Alcázar could be possible participants in the research. All visitors of the Royal Alcázar were free to join or to ignore the robot. However, not all visitors of the Royal Alcázar would enter the room during the study, as we did not run the study all day long, or visitors did not visit that part of the Royal Alcázar. We estimated that in both conditions together about 300 participants joined the robot; from shorter than one full explanation to an entire tour. People who stood close to the robot, but also people who followed the tour from a distance were counted in this number. Random people who joined the robot tour for at least one full explanation were chosen to participate in a short interview, the interview was with one person or with a few people who visited the Royal Alcázar and joined the robot together. When the robot performed human-like behavior, 6 interviews were recorded, while 11 interviews were recorded when the robot performed modality-specific behavior.

PROCEDURE

Before the robot started the guided tours, the researchers set up the cameras and placed the signs (stating that a scientific research was om progress and being recorded and that by entering the zone people gave consent to participate in the research and to use the recordings for analysis and presentations) at the entrances and at the tripods of the cameras. Other than previous performed experiments in the same room, we were not allowed to put signs to inform the visitors about the research on the floor.

The robot was brought into the room at:

- Tuesday 29-04-2014 at 11:00 to perform experiments until 12:00. During this session the robot performed guided tours in human-like mode.
- Wednesday 30-04-2014 at 11:45 to perform experiments. Due to some technical problems, the experiments were stopped at about 12:15. After fixing the problems we took another session from 17:00 till 17:50. During these two sessions the robot was performing the guided tours in robot specific mode.

When visitors entered the room during the performance of the study, they automatically gave consent to use the obtained video data for academic research and use of the material for publications, as they were warned by the signs (also, if one of the visitors wished not to be recorded, they could contact one of the researchers, recognizable by the FROG-tag and the film would be stopped or destroyed). Two clearly visible cameras were used to record the visitor reactions from two opposite corners of the room. Both cameras were placed at a height of 2 meters (maximum height on tripod). With the first camera the happenings in point 1 and 4 were recorded closeby. The second camera recorded the happenings close to that corner. For the analysis only the video recordings of camera 1 were used.

The tour given by the robot took about 3-5 minutes. When visitors entered in the Hall of Festivities, the robot was in the starting place (1) (see Figure 2) and started the tour by welcoming the visitors and giving some general information about the room. When the robot finished this story, it drove to the next stop, asking the visitors to follow. At the next stop (2) the robot told the visitors about the design of the figures on the wall, made with tiles, after which it drove to the next exhibit. At the third stop (3) the robot told the visitors about the banner that was hanging above an opened door. At the end of this story the robot asked the visitors to follow and it drove to the last stop (4) where it gave information about the faces visible on the wall. Before ending the tour the robot drove back to the starting point, informed the visitors the tour had finished and wished them a nice day. After a while, when new visitors had entered the room, the robot started the tour again.

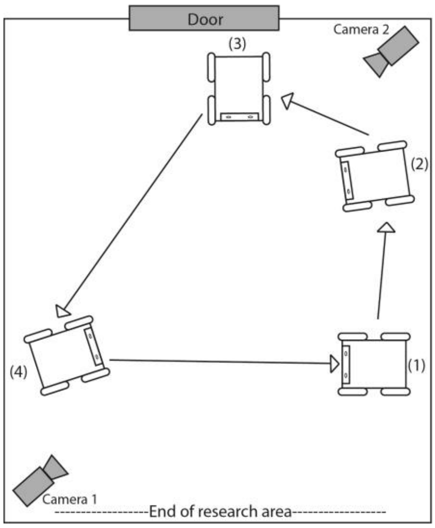


FIGURE 4: SCHEMATIC VIEW OF THE TOUR IN THE HALL OF FESTIVITIES

For the short robot tours, a timeline is given (same for both conditions), however, the timeline may differ per tour, as the time needed for getting to the next exhibit may differ due to people standing in front of the robot.

Overall timeline (in minutes) of robot guiding:

0.00-0.10	Robot picks up small group of visitors
0.10-0.30	Robot welcomes visitors
0.30-0:40	Robot drives to first stop
0:40-1.00	Robot tells something about the design of the tiles
1.00-1:20	Robot drives to second stop
1.20-1:40	Robot tells something about the banner above the door
1:40-2:00	Robot drives to third stop
2:00-2:20	Robot tells something about the faces of Charles and Isabel

2:20-2:40	Robot drives to endpoint
2:40-2:50	Robot says goodbye to visitors
2:50-5.00	Short interview with some of the visitors

Some of the visitors who followed a guided tour were asked to participate in a short interview about the robot guiding them. The interviews had a very open set-up. We had some questions prepared, but most important was that visitors told about their experience of being guided by the robot and whether they understood the intentions of the robot. The interviews took approximately 2 minutes in time, because the visitors were visiting the Royal Alcázar for the building, and not for the experiment. The interviews were recorded using a voice recorder. The interviews were performed in English (or Dutch).

Interview questions:

- What was your first impression of the robot?
- Can you describe your experience of being guided by the robot describe as if you would describe it to your friends or family at home?
- What do you think of the behavior of the robot?
- If the Royal Alcázar decides to have more robots such as this one guiding people, what would you think of that? Do you have any suggestions for the robot?

As the study was performed in a real-life setting, with uninformed naïve visitors, sometimes we had to deviate a bit from the procedure. The robot had defined places for stops, however, sometimes the robot had to stop close to the defined place, because people were walking or standing in front of the robot. Another reason to deviate was when the robot lost the attention of all the people following the tour. Then it drove back to the starting place and started over again. If some visitors lost attention and left the tour but other visitors remained listening to the robot, the robot continued the tour.

At the end of each session, we cleaned the room of the Royal Alcázar, and took away the camera's, the signs and the robot. The UT people checked the recorded film-material and copied it from the camera to a laptop when necessary.

Data analysis

For the analysis of the video data, we used DREAM - a thin-slice approach to annotate and analyze HRI in-the-wild data. This is a method that we developed to analyze video data in a fast, focused and standardized way. We report on this method in a paper that we have submitted to CHI 2015 (Karreman, Ludden, and Evers 2015b). In this section we describe briefly the steps that we took before analyzing the data. Parts of this section are copied from the paper.

Step 1: Operationalization of the research question

Our main research question for the study was: What are the differences in the reactions of naïve visitors of the Royal Alcázar to the different FROG personalities during the explanations? From the research question, we knew we had to focus on the moments the robot offered information about an exhibit. Knowing this, we determined the level of detail we needed to code and analyze the data. To be able to yield a good impression of the reactions of the visitors during the story of the robot not all video data is needed; for example, we are now not interested in the moments that the robot drives from one stop to the next. The actions during the tour that we are interested in are the moments the robot offers information about an exhibit; which are clearly recognizable actions in the tour. Therefore, we decided to use a sequence of three images to analyze the visitors' reactions to the two different robot personalities. To create the sequence we decided to use stills of the beginning of an action (captured when the first word of the explanation was played), the middle of an action (captured at a specific word in the middle of the explanation) and the end of an action of the robot (captured when the last word was played).

Step 2: Create images and sequences

To prepare the data, sequences of three screenshots of each explanation were abstracted. These screenshots were taken at three moments during the human robot interaction at a Point of Interest, as described above. For the analysis of reactions of visitors around the robot tour guide we used the sequences of screenshots at each of the four exhibits. This led to 144 sequences for the two conditions together; 74 for the human-like condition and 70 for the modality specific condition.

Step 3: Create codes

All sequences of screenshots were printed, so that two researchers were able to easily discuss all sequences and compare them with other sequences to verify whether their interpretations seemed to be correct. Then the researchers started to physically cluster the sequences with the same kind of events or occurrences, while they sometimes had to discuss their interpretations to come to an agreement for a cluster or to create a new cluster. Together the two researchers checked all clusters for accidentally misplaced sequences. Also large clusters were subdivided into smaller clusters and small clusters were combined into larger clusters where possible. Subsequently, the clusters were given names that were descriptive, distinctive and intuitive. These descriptions of clusters were the codes to be used in the next step.

Step 4: Apply codes to the reduced data

Then, the codes were given to a third researcher to annotate a sample of 33 sequences (about 20% of the data using all defined codes). The annotator was informed and trained first in a discussion with the primary researcher. Thirty three of the 144 sequences (of the explanations at the exhibits) were double-coded by two researchers using the analysis program Atlas.ti (Friese 2014). These two coded sets were used to determine the inter-rater reliability. To calculate the inter-rater reliability the online program CAT (Coding Analysis Toolkit) was used (Lu and Shulman 2008). This program is recommended by the developers of Atlas.ti. Using

CAT we found an overall inter-rater reliability of 0.61 (Cohen's Kappa) between the two coders and their coded data sets. This number indicates a moderate agreement between the coders on reliability of the codes used. This number is not very high, but it indicates the inter coder agreement is high enough to assume that the codes and the definitions are reliable. Therefore, the main coder went on coding, but to improve the agreement between the coders, first the differences between the coders were examined by the main researcher to understand better the choices made.

After the inter-rater reliability check, one researcher (who was always one of the coders to calculate the inter-rater reliability) coded all data. The data was coded in Atlas.ti. For all codes that applied to a sequence, the code was linked to the middle picture of the sequence. Not only the recorded data, but also the interview data was coded using Atlas.ti to obtain a complete overview of visitor reactions and behavior to the different robot personalities.

Step 5: Analyze data

After applying the codes, we analyzed the data. In this case all codes were analyzed by using Atlas.ti. No observation notes of the sessions in the Royal Alcázar will be used, because the researcher did not take notes as she was busy doing interviews,. However, the researcher had a broader view of the situation than is visible on the recordings, so something remembered from the sessions can help the analysis or help to find links between observations.

The codes applied to the sequences were counted and differences between the two conditions are presented (see appendix 1).

Figure 5: example of sequence to be analyzed

Results

As the study was a real-world study, sometimes we deviated a bit from the given procedure. Therefore we present here the number of stops for each exhibit in both conditions. In the human-like condition, the robot made in total 74 stops to explain something to the visitors. In the modality specific condition, the robot made in total 70 stops.

Table 13: number of stops per exhibit per condition

	Human like condition	Modality specific condition
Welcome	14	12
Tiles	17	17
Banner	9	17
Faces	17	15
End	12	9
Wrong explanation	5	0

Totals	*74*	*70*

This table shows that the robot made approximately the same number of stops per exhibit.

In the next section, the description of the results below, we will give numbers between brackets, for example (5), which indicate the number of specific occurrences happening.

H1: People will react differently to the two robot behaviors

We assumed that people would react differently to a robot with human-like behavior than to a robot with modality-specific behavior, for example in their way of listening to the robot and following the robot, their gestures to the robot, their formation around the robot, whether they take pictures and pose with the robot, as well as in their answers to interview questions. To test this hypothesis, we counted the occurrences of all codes in both conditions and checked these for large differences. See appendix 1 for all codes and the number of times they occurred. Also, we analyzed the differences in the answers to the interview questions between both conditions. As we found some differences in the behavior of the naïve visitors during the robot tours between the different conditions, we can adopt this hypothesis.

In the human-like condition none of the visitors present showed interest in the robot (16) more often than in the modality specific condition (2). Also, we found that more people in the modality specific condition (18) took a picture of the robot than people in the human-like condition (5). In the modality specific condition just one person was engaged with the robot (12) more frequently than in the human-like condition (5). From the observations we assume this is because the robot in modality specific condition explicitly asked people to stand closer, and sometimes one person did this. This explanation is strengthened by the fact that in the modality specific condition one person is standing very close to the robot (11) more often than in the human-like condition (1). In the modality specific condition, more often one person got disengaged (10) than in the human-like condition (6), while in the human-like condition two people disengaged at the same time (17) more often than in the modality specific condition (8). It was notable that in human-like condition fewer small groups – 3-5 people – disengaged at the same time (9) than in the modality specific condition (15).

Only people in the modality specific condition (3) mentioned that the voice of the robot was unnatural and mechanical, while one person in the human-like condition mentioned that the voice was clearly understandable. Therefore some people from the modality specific condition suggest the use of a more natural voice, while people from both conditions suggest the use of more different languages for the explanations. In general people in the modality specific condition gave more suggestions for improvements that could be made to the robot and the behavior of the robot. For example, a visitor should be able to influence the topics in the story told, and the robot should have its own space in the Royal Alcázar, as well as the remarks on the mechanical voice.

H2: People will prefer the robot with the modality specific personality over a robot with human-like behavior

Our second hypothesis stated that people would prefer the robot with modality specific behavior over the robot with human-like behavior, because it makes its intentions more clear. We expected that from the interviews we would find that people preferred the robot with the modality specific personality as a robot tour guide. To test this hypothesis we analyzed the interviews and searched for differences in the answers of people in the human-like condition and in the modality specific condition. However, from the interviews we did not find clear indications to adopt this hypothesis. People from both conditions used positive descriptions for their experience and described the robot as fun, interesting and different. The findings from the interviews as reported under hypothesis 1 suggest that people in both conditions experienced the robot in a different way. However, from these findings we did not find a preference for one of both conditions. Therefore, we have to reject this hypothesis.

However, we found that six people in the modality specific condition answered positive to the question whether there should be more robot guides in the Royal Alcázar, against one person in the human-like condition. Also, more people in the human-like condition (2) than in the modality specific condition (1) would prefer a human tour guide, while five people in the modality specific condition mentioned that they really do not like tour guides and the large groups, against none of the participants in the human-like condition. One person from the human-like condition thought that a robot tour guide would be a good alternative to the human tour guides. Four people in the modality specific condition mentioned that they liked the fact that they could leave the robot when they wanted to go somewhere else or when they wanted to focus on something else, without offending the guide that they left the tour. None of the people in the human-like condition made this remark.

H3: People will react to the robot with the modality specific personality as they would do to a human tour guide, for example in orientation and looking at the exhibit

We assumed that people would react to the robot with the modality specific personality as they would do to a human tour guide more than they would react to a robot with human-like behavior, because the one with modality specific behavior makes it intentions more clear. To test this hypothesis we counted the occurrences of all codes that had to do with the formation of the visitors around the robot in both conditions and checked these for large differences. Other than expected based on the results of the previous hypothesis, we have to reject this hypothesis.

Between the formation the visitors made around the robot, we did not find major differences. In both conditions people stood in line formation (human-like condition: 2; modality specific condition: 6) and in a semi-circle formation (human-like condition: 14; modality specific condition: 16), which we would also expect to find with human tour guides. Also, in both conditions people stood in a far away formation (human-like condition: 13; modality specific condition: 19) and there are in both conditions often more than one person/group at a time in listening to the robot formation (human-like condition: 25; modality specific condition: 24), which we do not expect to find with human tour guides. We conclude from this that the findings for both conditions are comparable and as the findings above described do not indicate differences between both conditions and as the reactions of visitors in both conditions deviate greatly from the reactions of visitors to human tour guides, we have to reject this hypothesis.

H4: People are better able to keep their attention to the story when guided by the robot with modality specific behavior

This hypothesis stated that people in the modality specific condition would have the ability to pay attention to the robot, the exhibits and the story of the robot for a longer time. To test this hypothesis, we counted the occurrences of all codes about the number of stops people followed the robot in both conditions and checked these for large differences. We assumed that people are better able to keep their attention on the story of the robot when guided by the robot with modality specific behavior and due that they would follow the robot for more stops. Based on the results we can partly adopt the hypothesis.

We found no major differences between following three (human-like condition: 8; modality specific condition: 6), four (human-like condition: 5; modality specific condition: 8) or five (human-like condition: 2; modality specific condition: 2) stops. However, we found a difference for following one and two stops between the two conditions. More people followed the robot in the modality specific condition for two stops (20), than that people followed the robot with human-like behavior for two stops (11). However in the human-like condition more people walked actively towards the robot (23) than people did in the modality specific condition (14). People in the human-like condition tended to follow the robot more often for only one stop (25) than people in the modality specific condition (18).

Results not related to a hypothesis:

In total 17 interviews were performed; 6 interviews were held with people who observed the robot in human-like condition and 11 interviews were held with people that observed the robot in modality specific behavior. It was noticeable that most people who observed the robot with modality specific behavior gave more extensive answers to the questions and allowed us to hold longer interviews, because they were willing to tell a lot about their experiences.

Participants from both conditions indicated that they liked the information given (in English) because it was information that they would otherwise not have obtained, because in the different rooms of the Royal Alcázar not much information is presented on the information boards.

Common in these interviews was that participants in both conditions told that the robot would appeal to children. In some cases there were children around following the robot and they only had attention for the robot, which then became an example of how things could work. Also, participants think that when children follow the robot they will pay more attention to the exhibits as well, therefore it could be a good teaching tool.

Discussion and Conclusion

This study was set up to evaluate the reactions of the visitors to find whether a human-like personality or a modality specific personality would better fit a tour guide robot. Just as in the online study, the sets of behaviors in this study had only subtle differences. Still, the reactions of the visitors differed clearly between the two conditions, which proves us that we were able to create two different robot personalities. The results of the interviews in this study do not show clear preferences for one specific robot guide behavior, but we did find that the modality specific personality seemed to attract our target group more.

We found that people do react differently to the two behavior sets of the robot, but that none of the visitor reactions to one of the robot behaviors closely resembled the visitors' reactions to human tour guides. We would assume that people would react the same in both conditions, because we copied human tour guide behavior and designed the behavior for the modality specific personality to realize the same interactional outcomes as human tour guides realize. The fact that the reactions differ from the reactions to a human tour guide makes clear that a robot is not a human and therefore also behaviors should differ. Hence, we did realize the interactional outcomes in the modality specific mode, as we were able to draw peoples' attention, direct peoples' attention to the exhibits, and guide them to the next Point of Interest.

We could not find a preference for one specific robot personality from the participants' answers to the interviews, because this preference seems to be very personal, as we also concluded from our online study. Noticeably, we found that in the human-like condition the participants of the interview stated they preferred human tour guides, while in the modality specific condition participants stated they did not like guided tours and large groups. These results might indicate that the different robot personalities attract different kinds of visitors. The robot with the human-like behavior seemed to attract people that liked human tour guides, while the robot with modality specific behavior seemed to attract the people who did not like human tour guides. A possible explanation for this might be that people who like tour guides recognize the behavior of the robot with the human-like behavior and therefore decide to join the tour, while the people who do not like tour guides decide not to join that tour. And vice versa, people who do not like tour guides, were attracted by the very different behavior of the robot with modality specific behavior. This is an interesting finding, because with the FROG robot we do not want to attract all people in a tourist site. Our target group (small groups that do not like to follow tour guides) seemed to prefer the robot with modality specific behavior, and therefore a guide robot with modality specific personality should be used in tourist sites.

We found that in the human-like condition more often none of the visitors present were interested in the robot in the modality specific condition, which might indicate that the robot with human-like behavior was more often ignored by the visitors. We have to note here that 3 times was coded that the robot in the human-like condition was close to or in a large group accompanied by a human tour guide, and the guides asked for attention and did not want the visitors to pay attention to the robot, while the robot in modality specific behavior was never in such a large guide group. Also, in the human-like condition the robot was slightly more often not visible (11) because a group blocked the view or the robot was out of the angle of view of the camera, than in the modality specific condition (8). However these numbers are close to each other, therefore, we can still conclude that people in the modality specific condition paid more attention to the robot. Connected to this finding, we found that people in the modality specific condition took many more pictures of the robot. This might indicate that the robot with modality specific behavior was significantly different from what people are used to, so that they wanted to have a souvenir (a picture) to remember and probably tell to friends about the robot later. We have to note here that for the analysis of taking pictures only screenshots were used, which might indicate that we missed people taking pictures at other moments. However, an explanation just took about 20 seconds, and we had three pictures to analyze for this time span. When people took a picture it was most often visible in at least two of the pictures of the sequence. Therefore, we assume we did not miss this even at other moments and that people in the modality specific condition took more pictures of the robot.

We found differences in the number of stops people followed between the conditions. People in the human-like condition followed just one stop more often than people in the modality specific condition, while people in the modality specific condition followed two stops more often. So, there is a difference in following for one and two stops between both conditions, but not for the other numbers of stops. For people who followed an extensive part of the tour (more than two stops), the difference in robot behavior seemed not to have any influence, because they seem to be interested in the story. The robot with the modality specific behavior seemed better able to attract and keep the attention of visitors for more than one stop than the robot with human-like behavior did, which indicates that the robot with modality specific behavior is better in keeping the attention of the visitors. These results might indicate that the robot with human-like behavior is better able to catch the attention of the visitors, while the robot with modality specific behavior is better able to keep the attention of the visitors. Also, the fact that people in the modality specific behavior gave more suggestions, might suggest that these people had more focus on the behavior of the robot, because it drew attention as it was different from what they were used to, other than the behavior shown by the robot in human-like condition. This might be due to a novelty effect, but we assume that it is a result of the fact that the robot in modality specific condition is better able to clarify its expectations of the visitors by giving clear instructions, which we have also described in the results of hypothesis 2. There we learnt from the interviews that when the participants were asked if they understood the behavior of the robot, from both conditions they answered that they understood, but only people in the modality specific condition added to the answer that they had to follow the instructions and follow the robot. Individuals in the modality specific condition especially followed the instructions to come closer at the beginning of an explanation by the robot. For duos or small groups, this would be more difficult, because when one person stands close, the view of the others is blocked. People in the human-like condition did not receive these instructions, and therefore probably showed more awareness of other people around them. From the disengagement from the robot in the two conditions we presumed that the robot with human-like behavior mainly attracted pairs/couples that visited together, while the robot with modality specific behavior mainly attracted single persons or small groups.

We have three limitations in this study that we want to discuss. The first limitation already occurred during the study. The robot was always steered by the same person who practiced the two sets of behavior prior to the study. The differences between the two sets of behaviors were small, and because the person who controlled the robot remotely had to improvise sometimes, for example when people were standing in the place the robot was supposed to stand or when all people left the robot and new people were attracted soon after. This made the differences between both conditions even smaller. However, the conditions were still clearly distinguishable by the behavior of the eyes and the pointer as well as the actions in the screen. The body movement was not clearly distinctive anymore. However, as we found in our analysis, the different sets of behaviors still influenced the reactions of the visitors in different ways. Second, this study is based on observations and short interviews. Extensive questionnaires or long interviews would probably have given more insight into the thoughts of the people and ratings like we obtained in the online study, but we did not choose to perform questionnaires and long interviews. The people in the Royal Alcázar did not came there for the robot study, and did not want to spend too long evaluating the robot. Therefore, we decided to base this study on observations to complement the findings of the online study. The last limitation is in the analysis of the data of this study. Now, only information about the stops was used, no information of the periods that the robot was driving to the next exhibit, which might be interesting as well. For the development of FROG we chose to focus on the stops when designing the behavior and personality, because these are the moments that are crucial for a tour guide. Guiding is important, because during these moments the robot might lose visitors. However, if the robot cannot present the information at an exhibit in the right way, it would not be an effective tour guide. Therefore, we focused in this study and in the other studies on the robot personality and behavior especially at the stops of a tour.

In general we can conclude that we succeeded in creating two different personalities for the robot, because we found that people reacted in different ways to the different sets of behavior. People who do not like human tour guides seemed to like the robot with modality specific behavior a lot, because this behavior did not remind them of a human tour guide. However, some of the instructions given by the robot with modality specific behavior were not chosen well, for example, when one person is already standing very close, the robot should not ask people to come closer. By giving this instruction, one person might exclude others from the tour. In other cases, participants stated that they understood they had to follow the instructions of the robot. The robot with human-like personality seemed to catch the attention of the visitors more easily, probably because the behavior is already known and familiar, but people also lost their attention for human-like condition sooner. The robot in modality specific condition seems to be better able to keep the attention of the visitors for at least two stops.

Guidelines for robot personality

Just as in the online study, we found in the in-the-wild study people distinguished between the two robot personalities; in the previously described study people reacted differently to both robot personalities. Other than in the online study, we would now suggest to design the personality for a tour guide robot as a modality specific personality, because this personality was preferred by people who do not like human tour guides for a long tour, but like to obtain information, which is exactly our target group. Also, when the behavior of the robot resembles human-like tour guide behavior too much, people tend to prefer a human tour guide, while the robot with modality specific personality seemed to attract visitors who do not like tour guides when it showed modality specific behavior. This might be because the robot is a product and people expect a lot of a robot that tries to be as human-like as possible in this case in the behavior.

In this study the appearance of the robot did not seem to have as much influence on the visitors' perception of personality as in the online study, because we found clear differences in people's reactions to the two different robot personalities. However, several people made remarks about the modern appearance of the robot, that does not fit the old buildings of the Royal Alcázar that well. The question is whether a more traditional or more human-like appearance for the robot would suit the function of the robot in this case. We expect that the modern appearance of the robot explicitly communicates its function and functionality.

Furthermore, the guide robot with human-like personality did not give explicit instructions, because in human-human communication, lots of the cues given in interaction are implicit, especially by gaze (Admoni et al. 2011; Broz et al. n.d.; Kleinke 1986). However, other than (Admoni et al. 2011), we found that for a tour guide robot it is effective to give instructions on the screen on what it expects the visitors to do as the robot did in modality specific condition, just as already was described by Clodic et al. in (Clodic et al. 2006), that the robot would lose the visitor interest when it does not communicate it status and intentions. Until date, robots cannot give all cues implicitly as human do, therefore explicit cues are supportive for the interaction.

Also in this study the visitors gave some suggestions for improvement of the robot. Just like in the online study, people prefer a (prerecorded) human voice for the robot, because otherwise it is difficult to listen to and understand. You might argue that this does not fit the idea of a more "machine-like" information point rather than a copy of a human tour guide. However, it seems that people do not relate the human voice with a copy of a human tour guide, while a mechanical voice is related to boring and difficult to understand.

APPENDIX 1

	Human-like condition	Modality specific condition
*No robot	11	8
*No visitors	2	4
*Robot in group of tour guide	3	0
AR_banner	9	17
AR_end	12	9
AR_faces	17	15
AR_tiles	17	17
AR_welcome	14	12
AR_wrong explanation	5	0
AV_active walk to robot	23	14
AV_look at object robot told about	5	4
AV_no clue where to look	0	1
AV_obstruct robot	1	0
AV_pose with robot	0	2
AV_short glance at robot	9	0
AV_take picture of robot	5	18
C_get interest in robot	4	3
C_observe robot	11	7
C_scared for robot	1	0
C_very close to robot	3	1
C_wave to robot	1	2
ENa_1 person engaged	5	12
ENa_2 people engaged	12	12
ENa_3-5 people engaged	25	25
ENa_6-11 people engaged	7	9
ENa_11+ people engaged	0	1
ENd_1 person disengaged	6	10
ENd_2 persons disengaged	17	8
ENd_3-5 persons disengaged	9	15
ENd_6-11 persons disengaged	3	1
ENn_1 new person atracted	10	10
ENn_2 new people attracted	7	5
ENn_3-5 new people attracted	15	13
ENn_6-11 new people attracted	1	0
F_1 person as close as possible	1	11
F_2 people as close as possible	7	6
F_2-2+ groups	25	24
F_closed group not interested in robot	2	0
F_formation of group is unstructured	17	20
F_large "guide" group	2	0

F_standing far away	13	19
F_standing in a line	2	6
F_standing in a semi -circle	14	16
P_0% (none)	16	2
P_20% (minority)	9	8
P_40% (little less than half)	11	14
P_50% (exact half)	7	11
P_60% (little more than half)	10	10
P_80% (majority)	14	13
P_100% (all)	5	4
S_five stops	2	2
S_four stops	5	8
S_more than five stops	1	0
S_one stops	25	18
S_three stops	8	6
S_two stops	11	20
T_1-3 people	5	7
T_4-10 people	39	36
T_10+ people	28	19

7. Overview of guidelines for FROG personality

From both studies we formulated some guidelines to design personality for the robot. In this chapter we will give an overview of all guidelines.

- A robot personality profile should be defined before making a visual appearance. Then the appearance and personality can be designed according to the profile.
- The behavior of the robot should not be a direct copy of human-like behavior, but designed especially for the modalities of the robot based on the interactional outcomes of human tour guides.
- The behavior of the robot should not resemble human-like tour guide behavior too closely, because the robot with modality specific behavior seemed to attract visitors who did not like tour guides when it showed modality specific-behavior. Also, the modality specific personality for the robot attracted visitors more often and for a longer time span.
- It is good if the robot gives instructions on the screen on what it expects the visitors to do as the robot did in the modality specific condition.
- The robot voice should preferably be a human voice, to ensure people will understand the robot easily.

8. How these guidelines affect work of other partners in the project

The design guidelines for the robot personality as given in the previous section, will be used to inform the other partners about the envisioned robot personality. The other partners will benefit from using the defined personality for the FROG robot, because it will help them to make the robot behavior consistent.

Design

Design and personality should match. The design of the robot is already finished, therefore the personality should match the design. If necessary only minor changes can be made. In the online study we found that people perceived the robot (for both designed personalities combined, as the scores were very close) as very interesting and honest (scores above 3.7), and cheerful, relaxed, easy-going, cute, lively and modest (all had a score above 3.4). The robot was not really perceived as provocative, obtrusive, untidy, silly or boring (all scores below 2.8). Based on these results, we advise to make some small changes to characteristics of the robot appearance, to achieve a robot appearance that is also perceived as more easy-going, dominant and serious and a bit less modest than it is now.

Content

The robot itself will not talk a lot, it will only have prerecorded speech for short sentences; the voice of the robot will be mechanical. From previous research we know that people tend to talk to robots when they have the idea that the robot is able to talk to them and can understand them. The FROG robot does not have ears, and only will use prerecorded speech, therefore we assume that people will understand that the robot will not hear or understand them during the tour.

However, during the tour the robot needs to give information about the exhibits to the visitors. This will be done with multimodal interaction. However, most of the time speech is necessary to explain things. To solve the problem of having a non-talking robot, we use a prerecorded voice-over to give information. From the real-world study as well as from the online study, we learned that people did not like the mechanical voice and would prefer a prerecorded human voice.

Also, from the real-life study we learned that the instructions given on the screen were effective to communicate the robot's expectation on how the visitors should react. Therefore, we advise to instruct people and give feedback on the actual status of the robot in a similar way.

Navigation

The way the robot navigates and approaches people and objects is part of its behavior. In deliverable 4.1c we give guidelines for approach behavior and drive behavior. These behaviors should be consistent over time to help define the personality.

9. DISCUSSION

As we have discussed the findings of the separate studies in each chapter, in this chapter we want to answer and discuss our research questions based on both studies performed. Both studies gave us different results; quantitative as well as qualitative and we had to compare and combine these results to be able to answer the research questions. Both studies also had their limitations as discussed in the discussion sections in chapters 5 and 6. However, when keeping the limitations and differences in mind, the studies led to several insights.

Can and how does the behavior of the robot affect the personality that people attribute to the robot?

To answer this question, we can only use the results of the online study, because only these reflect the thoughts of the visitors. The behavior of the robot does influence the personality that people attribute to the robot, however it only influences the perceived robot personality for a small part. The appearance of the robot defines the majority of the robot personality. Hence, it is important that the appearance and the behavior match, otherwise interaction with the robot will be perceived as strange or unfamiliar and therefore distracting.

Can and how does the behavior of the robot affect how human-like or machine-like people think the robot is?

The behavior of the robot only affects the personality of the robot very subtly. We found in the online study that the robot with the human-like personality was perceived to be more mechanical (than organic) and more serious than the robot with the modality specific personality. Participants did not distinguish between a human-like robot personality and a modality specific robot personality. However, about half of the participants stated they observed a difference between both robot personalities.

Also, in the in-the-wild study we observed that people reacted differently to the robot. Noticeable was that participants of the interviews in the human-like condition stated they would prefer a human tour guide more often than participants in the modality specific condition, while participants in the modality specific condition stated more often they did not like human tour guides. They also mentioned that they liked the robot as a means of receiving information. This might indicate that the robot in human-like condition was perceived as more human, because people recognized but did not like the copy of the human behavior applied to the robot. That made it inferior to a guide. While the robot in modality specific condition was seen as another class; a mechanical information point. People who stated to not like human tour guides preferred the robot to obtain information from but also indicated they liked that they could leave and return whenever they wanted without offending the robot. This might indicate that the robot with modality specific behavior was perceived to be more machine-like.

Do people prefer either a human-like personality or a modality specific personality for a tour guide robot, and, why?

The preference for one of both robot personalities seems to be very personal and there is no overall preference for one of both robots. About half of the participants of the online study liked the robot with human-like personality better, while the other half liked the robot with modality specific personality better. The order of the videos did not have any influence on the participants' preferences. In the results of the in-the-wild study we also did not find a preference for one of both robot personalities. However, the fact that participants in the human-like condition preferred a human tour guide, while people in the modality specific condition preferred the robot with modality specific behavior over a human tour guide might indicate that our target group (small groups of people who will not join a human tour guide but do like to receive some information about the site) would prefer a robot with a modality specific personality.

Are people better able to pay attention to the art works when guided by a robot with modality specific behavior, and what causes the effect?

Based on the results of both studies we would answer this question as undecided. However, when we put the emphasis on the in-the-wild study we can state that the tendency is that people can keep their attention on the robot with modality specific behavior better. Based on the online study, the answer would be “no.” Participants were more distracted when they saw the robot with modality specific behavior than when they saw the robot with human-like personality. This might be because the behavior of the robot with modality specific personality was unfamiliar to them and they needed to pay attention to the robot behavior to understand the actions of the robot. The robot with human-like personality showed more familiar behavior and distracted the participants less from the content. Finally, people will experience the robot as they did in the real-world study, therefore the results of this study are weighted heavier to answer this question. In the in-the-wild study we found that the robot with human-like personality attracted people's attention more easily, but the people also lost their attention very soon. People left the robot in the human-like condition after one stop more often than people in the modality specific condition. In the modality specific condition people followed the robot for two stops more often than in the human-like condition. This indicates that people are better able to pay attention and probably found the robot more fun for a longer time to the robot with modality specific personality and the story that the robot told. No differences were found between the two conditions when people followed more than two stops, which indicates that the effects of the personality on the longer tours is negligible.

The answers to these four research questions and the proposed design guidelines as presented in chapter 7 form the basis for the development of a personality for a tour guide robot. Overall we can conclude that we were able to create two personalities that were perceived different by the participants, even though we did base these personalities on the Product Personality Scale and used a product design approach to design the personality.

10. Conclusion and Future Work

In our studies on robot personality for a guide robot we used the Product Personality Scale and an industrial design approach rather than the Big Five or the MBTI personality inventory and dimensions and a behavioral informed approach, which are often used in human-robot interaction. We agree that personality for robots should be used to create a consistent set of behaviors. By using the Product Personality Scale and design approach we proved that we were able to develop two different robot personalities that people could distinguish from each other and that people reacted differently to, without using human personality scales.

We argue that the robot is a designed product and that a product needs designed behavior. A robot is not a human and we found that people reacted differently to the guide robot than to a human tour guide. However, the robot can offer other possibilities in a tourist site, such as an information point that people can join when they like and as long as they like. Therefore, we argue that a tour guide robot should not be designed as a copy of a human. In the online study we tested the two designed personalities, human-like and modality specific, and indeed we found that a modality specific was perceived as more organic and natural for the robot. But even more important, in interaction with the tour guide robot in the Royal Alcázar, people who not like tour guides liked to receive information from the robot with modality specific personality, while people in the human-like condition who do like tour guides did not like the robot more than a tour guide. As our target group seemed to prefer the robot with modality specific personality, we argue that designing appearance and behavior for a tour guide robot should be approached in a product design manner keeping in mind the modalities of the robot as well as the wishes of the target group.

With this information we want to go on with the development of the robot personality and behavior for tour guide robots. A next step that would be interesting for FROG is to evaluate the effectiveness of the different behaviors of the robot during the transfer from one stop to the next. Until now, the focus was not on this part of the behavior of the robot, but the data is available and it would add a lot of knowledge to effective guide robot behavior and personality. A next step for developing the personality and behavior of the robot would be to create more robotic behavior and personality for the verbal behaviors of the robot, as was done for gestures and movements during this study. A further step, that goes beyond the FROG project, is to find the level to which robot appearance and behaviors can be abstracted while they are still meaningful and understandable.

REFERENCES

Admoni, Henny, Caroline Bank, Joshua Tan, Mariya Toneva, and Brian Scassellati. 2011. "Robot Gaze Does Not Reflexively Cue Human Attention." Pp. 1983–88 in *Proceedings of the 33rd Annual conference of the Cognitive Science Society*.

Ajzen, Icek. 2005. *Attitudes, Personality, and Behavior*. McGraw-Hill International.

Broz, Frank, Hatice Kose-bagci, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. n.d. "Investigating Mutual Gaze." 3–4.

Burgard, Wolfram et al. 1999. "Experiences with an Interactive Museum Tour-Guide Robot." *Artificial Intelligence* 114(1-2):3–55. Retrieved February 9, 2012 (http://www.sciencedirect.com/science/article/pii/S0004370299000703).

Clodic, Aurélie et al. 2006. "Rackham: An Interactive Robot-Guide." Pp. 502–9 in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on Robot and Human Interactive Communication*. Hatfield, UK: IEEE. Retrieved (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4107857&tag=1).

Desmet, Pieter M. a., Juan Carlos Ortíz Nicolás, and Jan P. Schoormans. 2008. "Product Personality in Physical Interaction." *Design Studies* 29(5):458–77. Retrieved June 10, 2014 (http://linkinghub.elsevier.com/retrieve/pii/S0142694X08000525).

Digman, J. M. 1990. "Personality Structure: Emergence of the Five-Factor Model." *Annual Review of Psychology* 41(1):417–40. Retrieved (http://www.annualreviews.org/doi/abs/10.1146/annurev.ps.41.020190.002221).

Duffy, Brian R. 2003. "Anthropomorphism and the Social Robot." *Robotics and Autonomous Systems* 42(3-4):177–90. Retrieved March 13, 2012 (http://linkinghub.elsevier.com/retrieve/pii/S0921889002003743).

Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. "A Survey of Socially Interactive Robots." *Robotics and Autonomous Systems* 42(3-4):143–66. Retrieved July 12, 2012 (http://linkinghub.elsevier.com/retrieve/pii/S092188900200372X).

Friese, Susanne. 2014. *Qualitative Data Analysis with ATLAS.ti*. 2nd editio. Sage.

Gockley, Rachel, Jodi Forlizzi, Reid Simmons, J. Social, and Behavioral Sciences. 2006. "Interactions with a Moody Robot." Pp. 186–93 in *HRI '06 - Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*.

Goetz, J., S. Kiesler, and A. Powers. 2003. "Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation." Pp. 55–60 in *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.* Ieee. Retrieved (http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1251796).

Goldberg, Lewis R. 1999. "A Broad-Bandwidth, Public Domain, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models." *Personality psychology in Europe* 7:7–28.

Graf, Birgit, and Oliver Barth. 2002. "Entertainment Robotics: Examples, Key Technologies and Perspectives." *Safety* 6(6):8–12.

Karreman, D. E., G. U. Sepúlveda Bradford, E. M. A. G. van Dijk, M. Lohse, and V. Evers. 2013. "Picking Favorites: The Influence of Robot Eye-Gaze on Interactions with Multiple Users." Pp. 123–28 in *In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Tokyo, Japan.

Karreman, Daphne E., Geke D. S. Ludden, and Vanessa Evers. 2015a. “Beyond R2D2: Exploring Approaches to Design Behavior for Non-Humanoid Robots.” P. SUDMITTED in *Proceedings of the 33nd annual ACM conference on Human factors in computing systems.*

Karreman, Daphne E., Geke D. S. Ludden, and Vanessa Evers. 2015b. “DREAM: A Thin-Slice Approach to Annotate and Analyze HRI In-The-Wild Data.” P. SUBMITTED in *HRI ’15 Proceedings of the 10th ACM/IEEE international conference on Human-robot interaction*.

Kidd, C. D., W. Taggart, and S. Turkle. 2006. “A Sociable Robot to Encourage Social Interaction among the Elderly.” Pp. 3972–76 in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.* Ieee. Retrieved (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1642311&tag=1).

Kleinke, C. L. 1986. “Gaze and Eye Contact: A Reseach Review.” *Pshychological Bulletin* 100(1):78–100.

Kobayashi, Yoshinori et al. 2010. “Choosing Answerers by Observing Gaze Responses for Museum Guide Robots.” *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* 109–10. Retrieved (http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5453240).

Kwak, Sonya S., Heeyoung Kim, and Myungsuk Kim. 2008. “Personality Design of Sociable Robots by Control of Gesture Design Factors.” Pp. 494–99 in *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. Ieee. Retrieved (http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4600715).

Kwak, Sonya S., and Myung Suk Kim. 2005. “USER PREFERENCES FOR PERSONALITIES OF ENTERTAINMENT ROBOTS ACCORDING TO THE USERS ’ PSYCHOLOGICAL TYPES.” *Japanese Society for Science of Design* 52(4):47–52.

Lee, Kwan Min, Wei Peng, Seung-A. Jin, and Chang Yan. 2006. “Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human?Robot Interaction.” *Journal of Communication* 56(4):754–72. Retrieved October 24, 2012 (http://doi.wiley.com/10.1111/j.1460-2466.2006.00318.x).

Lippa, R. A., and J. K. Dietz. 2000. “The Relation of Gender, Personality, and Intelligence to Judges’ Accuracy in Judging Strangers' Personality from Brief Video Segments.” *Journal of Nonverbal Behavior* 24(1):25–43. Retrieved June 4, 2014 (http://link.springer.com/article/10.1023/A:1006610805385).

Lu, Chi-Jung, and Stuart W. Shulman. 2008. “CAT (Coding Analysis Toolkit).” http://cat.ucsur.pitt.edu/.

Ludden, Geke D. S., Hendrik N. J. Schifferstein, and Paul Hekkert. 2006. “Sensory Incongruity : Comparing Vision to Touch , Audition and Olfaction.” in *5th International Conference on Design and Emotion*, vol. 31. Göteborg, Sweden.

Mugge, Ruth, Pascalle C. M. Govers, and Jan P. L. Schoormans. 2009. “The Development and Testing of a Product Personality Scale.” *Design Studies* 30(3):287–302. Retrieved (http://dx.doi.org/10.1016/j.destud.2008.10.002).

Mutlu, Bilge, Jodi Forlizzi, and Jessica Hodgins. 2006. “A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior.” Pp. 518–23 in *Proceedings of the IEEE-RAS Conference on Humanoid Robots (Humanoids 2006)*. Retrieved (http://pages.cs.wisc.edu/~bilge/pubs/MFH06.pdf).

Nourbakhsh, Illah R., Clay Kunz, and Thomas Willeke. 2003. “The Mobot Museum Robot Installations: A Five Year Experiment.” Pp. 3636–41 in *2003 IEEE/RJS International Conference on Intelligent Robots and Systems*. Las Vegas, Nevada: IEEE. Retrieved (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1249720).

Oh, Hyojung, Sonya S. Kwak, and Myung-Suk Kim. 2010. “Application of Unexpectedness to the Behavioral Design of an Entertainment Robot.” *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction - HRI ’10* 119. Retrieved (http://portal.acm.org/citation.cfm?doid=1734454.1734501).

Tapus, Adriana, and M. Mataric. 2008. “Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance.” *AAAI Spring*. Retrieved October 25, 2012 (http://www.aaai.org/Papers/Symposia/Spring/2008/SS-08-04/SS08-04-021.pdf).

Tapus, Adriana, and Maja J. Matarić. 2006. “User Personality Matching with Hands-Off Robot for Post-Stroke Rehabilitation Therapy.” *INTELLIGENT SERVICE ROBOTICS* 1(2):169–83.

Thrun, Sebastian et al. 1999. “MINERVA: A Second-Generation Museum Tour-Guide Robot.” Pp. 1999–2005 in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, vol. 3. Detroit, Michigan: IEEE. Retrieved February 9, 2012 (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=770401).

Walters, Michael L., Dag S. Syrdal, Kerstin Dautenhahn, René te Boekhorst, and Kheng Lee Koay. 2007. “Avoiding the Uncanny Valley: Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion.” *Autonomous Robots* 24(2):159–78. Retrieved October 10, 2012 (http://www.springerlink.com/index/10.1007/s10514-007-9058-3).

Woods, S., K. Dautenhahn, C. Kaouri, R. Boekhorst, and K. L. Koay. 2005. “Is This Robot like Me? Links between Human and Robot Personality Traits.” Pp. 375–80 in *Humanoid Robots, 2005 5th IEEE-RAS International Conference on.*

Woods, Sarah N., Michael L. Walters, Kheng Lee Koay, and Kerstin Dautenhahn. 2006. “Methodological Issues in HRI : A Comparison of Live and Video- Based Methods in Robot to Human Approach Direction Trials Ofnlaresa Culd Be Pilot.” Pp. 51–58 in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*.

Yamazaki, Akiko et al. 2008. “Precision Timing in Human-Robot Interaction : Coordination of Head Movement and Utterance.” Pp. 131–39 in *CHI ’08 Proceedings of the SIGCHI conference on Human factors in computing systems*.