Conversational Fillers for Response Delay Amelioration in Child-Robot Interaction

Master's Thesis Cognitive Artificial Intelligence
45 ECTS

Noel R. Wigdor

Supervisor TNO
dr. Joachim de Greeff

First UU Assessor
prof. dr. John-Jules Meyer

Second UU Assessor
prof. dr. Marcel van Aken

2014
Universiteit Utrecht

Abstract

Conversation Fillers (CFs) such as "um", "hmm", and "ah" were tested alongside iconic pensive or acknowledging gestures for their effectiveness at mitigating the negative effects associated with unwanted anthropomorphic robot response delay. Employing CFs in interactions with nine- and ten-year-old children was found to be effective at improving perceived speediness, aliveness, humanness, and likability without decreasing perceptions of intelligence, trustworthiness, or autonomy. The results also show that an experimenter covertly crafting a robot's vocalized response has a slower heart rate and a higher heart rate variability, an indication of a lower stress level, when the robot is filling the associated delay with CFs than when not.

Acknowledgements

I wish to thank my supervisor prof. dr. John-Jules Meyer for his continual good-natured support and mentoring during the course of my research. I am equally thankful for the close guidance I received from Joachim de Greeff, without which I could not conducted this research. Many colleagues at TNO were also of great help, notably Mira Kaiser for her unwavering support and hard work at every step of the way, and Rosemarijn Looije and Mark Neerincx for their generous guidance. Finally I would like to dearly thank Bert Bierman, without whose sincere devotion to my work and success I could never have made it here.

None of this would have been possible without the creative, amiable, and supportive people surrounding me in Utrecht. I would like to thank my friends Delphine Robache, Hans Bronk, Lara Solms, Jacco Spek, and Aafke Fraaije for making my time in Utrecht as memorable and lovely as it was. And of course I would like to finally express my appreciation for the support shown by my family throughout the last two years abroad.

Thank you all very much, it was an incredible experience I could not have had without your support and friendship.

Contents

1 Introduction

There is an inconvenient truth at the center of much Human-Robot Interaction (HRI) research. Those interested in HRI are pursuing a dream of autonomous, intelligent robotic agents that will interact with humans in a social manner. The truth, however, is that it is unclear when, if ever, this pursuit will come to fruition. Our longing for such robots has been popularized in literature, film, and television, and appears time and time again in the introductions to papers in HRI."[Robots] will play an essential role in human society in the not-so-distant future" (Kanda et al., 2004); "The day is quickly approaching when we will see mobile robots around us in our everyday environments" (Glas et al., 2009); or even "Imagine the near future in which one of your two personal service robots compliments you on how great you look today" (Midden and Ham, 2012);. The dream has pushed us to create anthropomorphic robots equipped with a plethora of human-analogous abilities, including speech synthesizers, cameras, and microphones. Implicit, and at times explicit, in the creation of such robots is the assumption that the processes behind human or human-like cognition can not only in theory be reproduced in a computer, but that the nearness of such a goal justifies human-like robots today.

Whether one believes that we are close to the intelligent robots of our fantasies or not, researchers have marched ahead and asked intriguing questions about what the interaction between everyday people and such robots will or should look like. Employing anthropomorphic robots, experimenters have conducted research which often resembles certain early work in psychology. We have learned about ideal conversational style, the effects of touch, and good gazing behaviour (to name a few) in the context of HRI. Implicit (or explicit) in such work is a conviction on the part of its researchers that such questions are worth asking because robots that can perform such functions autonomously are simply a matter of time (how much time, however, is far from settled).

This work is quite often made possible through a trick of sorts. As robots will inevitably be endowed with human-like ability, the argument goes, it is benign to covertly involve a human in the loop to provide the cognitive ability that robots will (soon or someday) have. The technique, called the "Wizard of Oz" or "WoZ" method, does just that. Named after the popular Frank Baum book and its later adaptation into a film in 1939 (in which the "terrible wizard" is revealed to be no more than a device controlled by a man hidden by a screen), the technique involves the teleoperation of one or more of the robot's actions (Riek, 2012). In HRI research, the method is used at many levels of the "autonomy spectrum", including the human control of movement, speech, gestures, etcetera, often unbeknownst to the human participant (Riek, 2012).

Robot interactors (humans interacting with robots) not aware of today's limitations of artificial cognition rely on preconceptions and non-technical impressions to guide their understanding of the nature of the robots they encounter. Research has clearly indicated that in the case of anthropomorphic robots, which are specifically designed to be treated as human-like in their ability to some extent, people have expectations akin to the dream of HRI researchers: intelligence, intentionality, and autonomy in line with the analogy that a robot's hardware is designed to illicit (Fischer, 2006). These expectations can be subdivided into two groups: the expectations that the robot can autonomously meet at present (e.g. walking, nodding, object tracking) and those it can not (e.g. autonomous conversation participation, mental states, emotion). When the WoZ technique is employed to give the (false) impression that a robot possess abilities in the second group, an illusion is either created or upheld in the mind of a 'naive' robot interactor. This (false) sense of robot autonomy in its interactor has sometimes been referred to as the "Illusion of Agency" (Midden and Ham, 2012), (Aucouturier and Ikegami, 2009).

The Illusion of Agency or IoA can falter in the face of robot behaviour that defies the expectations it produces. WoZ techniques, while useful in creating and maintaining this illusion, often have limitations which can diminish the extent of the IoA. The focus of this research is on the negative effects on IoA arising from response delay during conversations with an anthropomorphic robot in a WoZ setup. In this scenario a robot's appearance (perhaps combined with other factors) suggests to a interactor that it can converse naturally in a human-like way, an ability bestowed (usually covertly) on the robot through remote control by a human operator. The expectations are most grandiose and sustained with children (Nalin et al., 2011), making them an ideal candidate for the less-than-ideal robots at researchers' disposal today. Children are thrilled to converse with an anthropomorphic robot and, once the WoZ has made clear the conversational ability of the robot, will readily adapt their IoA to match (Nalin et al., 2011; Ros et al., 2011). While foreseen dialog from the child can have 'canned' or preprepared responses on hand for smooth robot reaction, the experimenter is often forced to generate on-the-fly dialog for the robot, which often leads to response delay (Kanda et al., 2009).

Insofar as the IoA demands that the robot respond in a human-like fashion, response delay can be an important, and unfortunate, violation of a child's expectations. Relatedly, from the experimenter's point of view the unnatural delay caused by off-hand script generation during dialog is a point of stress whose duration is not accurately estimated (Glas et al., 2012). This situation leaves bare the need for a tool to prolong the acceptable response time during dialog while not negatively affecting the IoA in other ways. One such method is to employ 'conversational fillers', that is, human-like sounds, often accompanied by gestures, that act as utterances in a conversation but are not intended to carry semantic content from the utterer to the listener (Shiwa et al., 2008).

Conversational fillers (CFs) are typically non-lexical 'grunts' or utterances that play a metalinguistic role in human conversation (Bortfeld et al., 2001). Says Bortfield: "If a speaker takes a long time to produce an utterance, she risks losing her addressee's attention or her speaking turn; but if she rushes to produce one that is defective, she risks being misunderstood". CFs often serve to 'fill' this gap with a mix of non-lexical utterances ('ummm', 'huh', etc.), iconic gestures (scratching of the chin, redirecting gaze, etc.), and lexical items intended to recognize the pause without ceding the conversational turn ('good question', 'let me think', etc.).

While Bortfield identified the risks to response delays in human conversation, within the context of HRI, unacceptably long response delays in dialog, common in WoZ scenarios with (somewhat) open conversation, carries with it a further risk: the potential loss of "believability". The term "believable" in HRI refers, according to one HRI researcher, Ben Robins, to "Agents and robots whose behaviour and interaction make narrative sense to humans as social beings..." (Robins et al., 2005). Believable agents are ones who behave and interact in accordance to their interactor's illusion of life-like agency for that agent.

The believability of robots is widely considered to be necessary for successful social interaction with humans. MIT's Cynthia Breazeal describes the situation as follows: "...Researchers have suggested that in order to interact socially with humans, a software agent must be believable and life-like, must have behavioral consistency, and must have ways of expressing its internal states" (Breazeal and Scassellati, 1999). It is important to stress that the believability of a robot is not the same as its believability as a human-like entity. An interactor's IoA for anthropomorphic robots may be decidedly non-human-like in some or many respects.

CFs have been employed with success in sustaining believability in the face of delays caused by on the fly dialog generation in HRI (Kanda et al., 2009). However, despite CFs being successfully used to ameliorate the negative effects of robotic response delay, surprisingly little research has been done to study what spin-off effects their use has, positive or negative. While CFs may help with the believability of child-robot conversation, it is yet unknown what subtle or large effects their use can have on the rest of the child's illusion of agency.

The topic of the research presented here centres on this question. It seeks to at once confirm that conversational fillers are effective at mitigating unacceptably long robot response times, and also to explore what other effects are brought along with their use, primarily for the interactor but also for the WoZ herself. Do CFs improve upon a robot's perceived likeability, aliveness, agency, intelligence, and response time without making the agent seem less trustworthy, friendly, liked, competent, or alive-like? Does their employment reduce the perceived time pressure on a WoZ or allow her to better estimate the time elapsed during an unwanted pause?

The answer to these questions is important not just to human-robot interaction but also to the artificial intelligence project as a whole. An important step in creating socially acceptable intelligence is fostering a healthy perception of social robots. Although our artifacts display what could be called a *sort* of intelligence (Google can search millions of pages in a fraction of a second), artificial intelligence is often conflated with human-like social intelligence. Some argue, like the famous Alan Turing or the anti-realist Michael Dummett, that AI's aim is not necessarily to create human-like machines that, through their emergent behavior, have human-like qualities. Instead, they argue AI should aim to create, through whatever means necessary, machines that can be *believed* to be intelligent. Insofar as the goal of AI hinges not on AI modeled closely on human-like cognition but instead on a robot's believability, any method that can help in this endeavor helps in our pursuit of AI. Techniques such as conversational fillers may not simply be 'tricks' that help with our robots' shortcomings. If AI with human-like cognition and emergent, intelligent behavior is unrealistic, tools like CFs may well be the only road to believability for the sorts of intelligence our computers can offer us.

These considerations are central to the Cognitive AI program at the University of Utrecht, the program for which this thesis is prepared. The program takes as an aim the engagement of students in the philosophical debate of the limits of machine cognizers. The debate is a wedge at the heart of much discussion of the past, present, and future nature of AI. Are intelligent agents those that behave indistinguishably from human agents (as Alan Turing advocated, an issue discussed at length in the 'Background' section of this document) regardless of their internal design? Or is human (conscious) intelligence a functional property of a cognitive system, as David

Chalmers argues, that cannot be judged by a system's behavior alone?

The employment of CFs in support of more grandiose, but admittedly *illusory*, views of the richness of robots' internal states in the minds of child interactors could be seen as a commitment to the former, and an affront to the latter, view on AI as a research endeavor. Insofar as CFs are designed to support the *false* view that a robot is intelligent and/or has mental states, work in this field commits itself to the validity of pursuing AI as the *appearance* of intelligence. For some, the pursuit of believability of robots pretending to have rich, human-like mental states while openly admitting that these robots do not actually posses complex mental states would be seen at best as a diversion from the important work needed in AI and, at worst, the setting of a dangerous precedent for future robot design that promotes deception. We will return to this debate in the 'Background' section to follow.

In any event, the experimental work presented here aims to help understand what effects CFs have on the believability of anthropomorphic robots in children projecting a human-like illusion of agency. It builds upon the widely cited work by Shiwa and Kanda who compared the effects of delays of various durations with or without CFs (Shiwa et al., 2008). While user evaluation scores decreased with extended response delays, CFs were shown to improve the score at each step (they tested three-, five-, seven-, and nine-second delays). While the results are promising, little can be gleaned on believability or other important metrics from a simple one-to-seven scale of evaluation of the response delay.

1.1 Research Questions

The current work's primary aim is to elucidate the effects of CFs on a child's IoA, that is, the change in the understanding one has of a robot's nature and abilities, in WoZ scenarios with unwanted robot response delay. Detailed questionnaires and other measures focusing on an anthropomorphic robot's perceived response speed, humanness, intelligence, autonomy, comprehensibility, and likability were administered. As we will see, the work suggests that not only do CFs improve subjective evaluation (as Shiwa and Kanda reported), for children they also improve the perceived speed of response, increase perceived humanness, and make a robot more liked. It has also been shown that these improvements did not come at the cost of decreased perceptions of intelligence, trustworthiness, or autonomy.

A secondary but important question addressed by this research is whether CFs can be used to reduce WoZ stress during unwanted robot response delays or make the delays' duration easier to estimate. The methods used to shed light on these questions - heart rate and heart rate variability and length of typed responses during filled or unfilled pauses of equal length - are novel approaches that build upon investigations by Glas and Kanda (Glas et al., 2012).

We will begin with a survey of the background relevant to the new experimental work along with some tangential but related research. The experiment's methodology and results will then be detailed, followed by a discussion of the results. Some space will be left at the end for a general discussion, possible future research suggested by this research, and appendices.

2 Background

2.1 ALIZ-E and TNO

The work presented here was conducted as part of the ALIZ-e project, an European Union-funded research program focussing on prolonged child-robot interaction. ALIZ-E ("Adaptive Strategies for Sustainable Long-Term Social Interaction") aims to employ Nao robots to help diabetic children cope with their affliction. It is quite friendly and non-threatening in appearance and is particularly well suited for child-robot interaction (Nalin et al., 2011).

To reach the goal of using a Nao as a "'educator' or as a peer 'motivator'." (Belpaeme et al., 2012), research within ALIZ-e has focused on strategies for *sustained* interaction, that is, fruitful interaction across multiple session separated by days or months. The work can be viewed as posing two challenges: to develop activities and technology to facilitate the teaching of health-related information to children with metabolic disorders, and the development of strategies for interaction design to make these activities and technology effective over prolonged periods of time. The present work is intended to contribute to this goal by investigating how CFs can be employed by a Nao robot in building/sustaining an IoA that heightens children's engagement.

Research in support of long-term interaction has contributed to a number of fields over the program's four-year lifespan, including (but not limited to) questions concerning the effects of child-robot touch in bond reinforcement, interpretation of non-linguistic utterances (Read and Belpaeme, 2014), robot 'emotion' (Tielman et al., 2014a).

TNO, the Netherlands Organisation for Applied Scientific Research, is a partner in the ALIZ-e project based out of the Netherlands. TNO is focused on conducting research to "contribute to the competitiveness of companies and organizations, to the economy and to the quality of society as a whole" [http://www.tno.nl]. The experimental work presented presently was conducted under the supervision of Joachim de Greeff and Rosemarijn Looije, researchers at TNO and active contributors to the ALIZ-e project.

The importance of the present work arose from apparent shortcomings in the interaction dynamics with target users (children aged 8 to 11 years old) in the ALIZ-e project. The goals and methodology of the ALIZ-e project were clearly outlined in a 2013 paper from the TNO ALIZ-e team describing a pilot study describing their work with diabetic children (Blanson Henkemans et al., 2013). In this paper, their research aim was summarized as: "How can a personal robot contribute to children's perceived enjoyment, motivation, and knowledge of diabetes" (it should be clarified that the enjoyment to which they wished to contribute was for the interaction with the robot, not for diabetes!).

The authors go on to list psychological needs that children have that "affect intrinsic motivation in game play" that come from self-determination theory or SDT. "SDT is based on the idea that players of all types seek to satisfy particular psychological needs in the context of play" say the authors. One of the three needs listed is 'relatedness', that is, "[the feeling that emerges] when a person feels connected with others. More specifically SDT hypothesises [sic] that environments that support perceptions of social relatedness improve motivation, positively influencing learning behavior". Put more simply, increasing the sense that the child has that they are interacting with a social agent in a natural way should support the primary goal of ALIZ-e cited above.

The present work was conducted to boost the feeling of relatedness within the child-robot interaction at the heart of the ALIZ-e project. It is hypothesized that CFs will not only reduce boredom and improve engagement but also improve the feeling that the child has that it is interacting with a believable social agent in a human-like manner. As we will see the CFs succeeded at making the Nao seem much more alive in the eyes of the children (while at once performing their primary function of mitigating the negative effects of response delay). This function - what ALIZ-e researchers call "credibility" - is in fact a stated goal on the ALIZ-e agenda, as relayed in the paper in (Ros et al., 2011):

> Another element we found to be useful in making the interaction happen more naturally is the use of additional behaviors, not directly related to the experiment or interaction itself (and that

> of course will not influence its result), but that can be useful in support [sic] the *credibility* of the robot as something which is "living".

The authors go on to relay that their experience in the ALIZ-e project has confirmed existing research by Sherry Turkle et al in 2006 in which 8-13 year-olds readily anthropomorphized humanoid robots and responded that they do indeed find them "sort of alive" (Turkle et al., 2006). This theme will be returned to in a following section (2.0.6).

2.2 Wizard of Oz

Often times HRI researchers or robot developers want to explore human-robot interactions involving capabilities not yet possessed by their robot actor. Often times this accommodation is done by having a human covertly control some aspect or aspects of robot behaviour. This technique - called the "Wizard of Oz" or "WoZ" approach - tends to be justified by one of three related beliefs: that the future behaviour will eventually be made autonomous by those conducting the research (easy prototyping), in anticipation that the behaviour could, at some later date, be technologically possible (and desired) but is not yet so, or that the behaviour could be automated but need not be so for the sake of the research at hand (Maulsby et al., 1993; Riek, 2012).

The human-in-the-loop, or simply called the WoZ herself, has been used to control a plethora of robot actions. In a 2012 review of fifty-four experiments employing WoZ techniques, it was found that 72.2% had the WoZ to perform some variety of Natural Language Processing, for instance, "having the robot appropriately respond to things the user said to it, making utterances, etc.". 48.1% used a WoZ to trigger non-verbal utterances such as nodding or pointing (Riek, 2012).

Adopting the WoZ approach has clear advantages for those interested in developing robots or studying HRI. In the former case, developers are free to test robot behaviour before investing the time, energy, and money into a standalone product i.e. rapid prototyping (Maulsby et al., 1993). In the latter case, using a WoZ allows a HRI researcher to establish an IoA (see the Introduction above) for the sake of experimentation involving either some other already automated behavior (i.e. as an interaction support) or to explore the nature of human-robot interactions that are anticipated for the future.

Controversy The approach, despite its popularity and utility, has been criticized for a myriad of reasons since its inception (dating back, at least, to 1987 (Maulsby et al., 1993), (Fraser and Gilbert, 1991).

One moral issue central to the WoZ method is its nature as a form of participant deception. Many times the participant is never informed that the robot is in fact being (partially) controlled by a human. This raises concerns that the research is 'making a fool' of its participants and could be said to undermine the requisite 'informed consent', that is, the ethical standard that participants in studies not be misled as to the precise nature of the experimentation (at least at the experiments' conclusion). This concern has led the Engineering and Physical Sciences Research Council (EPSRC), a UK funding agency, to insist that participants in WoZ-employing experiments should always be able to "lift the curtain", that is, ascertain the exact nature of the humans' involvement in the robot's behaviour. From their website (June 11 2014):

> "The legal version of this rule was designed to say that although it is permissible and even sometimes desirable for a robot to sometimes give the impression of real intelligence, anyone who owns or interacts with a robot should be able to find out what it really is and perhaps what it was really manufactured to do"

Apart from misleading just the participants involved in HRI research, using a WoZ approach without transparency carries with it, according to some, the danger of setting unrealistic visions of robot ability in the eyes of the public as a whole (Fernaeus et al., 2009). This can further reinforce the unrealistic image of robot ability (and AI) held by laymen (or even fellow researchers), which in turn (apart from being simply inaccurate) may set expectations for robots impossibly high and hinder their adoption and evolution.

A more technical but serious issue with the WoZ technique is with enforcing WoZ error rates to be in line with the future, autonomous system's error rate (and diligently reporting what the WoZ error rate was during an experiment). Almost all computer systems involved with real-world information processing will have a less-than-ideal success rate and insofar as the WoZ controlled aspects of robot behaviour are meant to be later displaced by automated systems it is important to enforce an expected error rate in WoZ situations. However, this is not practiced with due diligence. 81.5% of the fifty-four studies surveyed by Riek et al. in 2012 did not constrain the accuracy of the WoZ behaviour to be in line with reasonable expectations of future autonomous replacements and only 3.7% report measuring WoZ error rate at all (Riek, 2012). It can be argued that the studied interactions arising from a near-perfect performance by a WoZ can not be said to be strictly analogous to the interactions that would arise when the error from the future autonomous implementations is introduced. The quality of said study should be understood as being somewhat undermined by such lack of consideration. Fraser et al. make this point clear in their discussion of using WoZ setups in speech input/output computer systems (natural language processing, as discussed above, is WoZ-controlled in 72.2% of the WoZ-utilizing experiments reviewed by Riek) :

> "A...less obvious precondition is that before the experiments are begun it should be possible to formulate a detailed specification of how the future system is expected to behave...This specification often needs to be more precise and more detailed than would normally be necessary just to build the computer system. For example, in a speech simulation the [WoZ] ideally needs to make recognition errors at the same rate and in the same was as the future system. However, while descriptions of speech understanding systems often specify error rates, they rarely indicate what kinds of errors are made in sufficient detail for the errors to be simulated.

A less obvious criticism of readily employing the WoZ technique in HRI is that doing so allows its practitioners to neglect the more difficult task of creating the systems that the WoZ simulates. The danger, says Breazeal, an HRI researcher writing in 2005, is not only that the hard, technical work is left undone or that a future system's error rate may be misestimated or overlooked, but that the very *nature* of the human-robot interaction may well be different in important ways once an autonomous system operating under practical, technological constraints is created. Says Breazeal:

> This is done for good reasons, but it misses the opportunity to investigate how to design autonomous robots that successfully mitigate errors that inevitably do arise in human-robot teamwork [due to] common performance limitations.

(Breazeal et al., 2005)

This leads to a final danger in relying on WoZ-infused experimentation. Researchers (in fact, just about everyone) have been notoriously wrong in their estimation of what will soon be made possible in the realm of AI and robotics research. A particularly colourful illustration comes from work done in part by Apple Computer's Human Interface Group in 1993. The paper - "Prototyping An Intelligent Agent Through Wizard of Oz" - stresses that the authors intended to WoZ to simulate robot abilities that were obviously implementable, that is, obviously realizable. "In some ways, the [agent] we tested was more stupid than one we would implement" they claimed. The human-controlled abilities - natural language processing, machine learning, and iconic gesture recognition and processing - are still far from full fruition eleven years after their work was undertaken. Special care needs to be taken when assuming that a WoZ-controlled behaviour will soon be made possible. This was succinctly summarized by Fernaeus in 2009:

> A danger that has been noted with this and related methods is that users (and even researchers) may get lured to believe that the step to take from a Wizard of Oz setup to achieve a fully autonomous version of the system is much smaller than is the actual case. This is especially relevant in research that does not intend to end up in a working system, but serve other important purposes, e.g. to learn about how people behave together with embodied interactive artefacts on a more general level.

(Fernaeus et al., 2009)

Despite these concerns, the WoZ technique is an invaluable tool in human-robot interaction. It allows for rapid, low-cost prototyping to guide robot design, and, when carefully used, the study of the inevitable interactions between humans and robots that will arise alongside robots' appearance in layman's lives.

The work currently presented abstains from commenting on the merits of WoZ design and instead takes its use as a given problem worthy of support. As mentioned previously, the present work aims to use conversational fillers to fill unwanted gaps in human-robot dialog in WoZ scenarios when, for instance, a teleoperator must type a novel response to unanticipated participant questioning. We now turn our attention to human CF use and into their previous use for respose-delay mitigation.

2.3 CF in Natural Language

Human CFs have been extensively studied in the field of linguistics. They are treated as rather peculiar forms of utterances because of their seeming paralinguistic role in conversation. CFs by definition are not intended to carry semantic content like typical, lexical utterances do. Despite this oddity, CFs (loosely equivalent to 'grunts', a form of 'disfluency' or 'non-lexical word') are used extensively in natural human language, occurring approximately every five seconds in American English (Ward, 2000).

They do, however, seem to serve a number of important purposes. It has been shown that (certain types of) information following a CF can be better recalled than when the information is presented without a filler such as 'um', or 'theeee...' . It is thought that in this sense CFs signal the upcoming arrival of important information (Pfeifer and Bickmore, 2009).

They have also been shown to be a signal of an utterer's uncertainty in the correctness of a response, a signal that listeners are indeed sensitive to. In a study by Williams in 1995, answers to questions proceded by 'um' or 'uh' were rated as less likely to be correct by listeners (a result repeated by (Bortfeld et al., 2001). The nature of the utterance also seems to be indicative of a speaker's certainty: Ward also showed that 'um' and 'am', compared to 'uh' and 'ah', generally seem to indicate more thought on the part of the utterer in the mind of the listener.(Ward, 2006).

CFs have also been hypothesized to be useful in slowing the rate of information transmission for better listener comprehension. Speakers may be sensitive to listener "information-uptake capabilities" and use fillers to slow the rate of information transmission in an attempt to boost comprehension and remembering (Ward, 2000).

Perhaps most relevant to CF use in HRI, CFs have also been shown to be effective in holding a speaker's turn during a conversation while an upcoming utterance is formulated (Bortfeld et al., 2001). Says Bortfield: "If a speaker takes a long time to produce an utterance, she risks losing her addressee's attention or her speaking turn; but if she rushes to produce one that is defective, she risks being misunderstood...". However, it also appears that CFs can act as a sign of trouble on the part of the utterer and a signal for the listener to help complete the utterer's sentence for them (Bortfeld et al., 2001).

The frequency of CFs in human-human dialog points to a need for their use in human-robot conversation for increased naturalness. Their multi-faceted utility indicates a sea of opportunity for their use in human-robot dialog for better information transference and robot response delay mitigation. We now turn our attention the way in which CFs have already been used in HRI.

2.4 CF in HRI

The importance and potential advantages of using CFs in human-robot dialog has not gone unnoticed by HRI researches. Focus has centered on their use for "buying time" for a delayed robot response, typically caused by complications related to a WoZ manually constructing a robot response. Often times the fillers were used without them being a focus of research, that is, as an assumed aid in making delayed robot response less irritating to interactors. For instance, in 2009 Kanda and Shiomi incorporated CFs into a robot that gave verbal advice to shoppers based on their inquiries. While their system was designed to be mostly autonomous, a WoZ was nonetheless needed for speech recognition and "to handle unexpected situations". Say the authors:

> ...since users might feel uncomfortable during slow responses or long pauses, robot response time is critical. To solve such problems, we implemented a conversational filler to buy time [29]. When the operator needs a few seconds, he/she executes a conversational filler behavior to notify listeners that the robot is going to respond soon.

(Kanda et al., 2009)

CF use in HRI has, however, been studied directly in a number of ways. The types of this research can be divided into two classes: the effects of CFs on the WoZ using them and the effect they have on the interaction from the participant's point of view. We start with the latter.

Perhaps the most widely influential study on the matter comes from Shiwa and Kanda out of Japan. In their 2008 work they compared participants' subjective evaluations of conversations with a robot in which its responses were delayed 3, 5, 7, and 9 seconds. Their primary interest was in " conversational fillers [used] by a robot to moderate user frustration toward the delayed response" at each time interval. To accomplish this, participants were asked to simply provide a one to seven evaluation with one standing for the lowest in a four-by-two design (four time intervals, two conditions). Shorter response times were filled using the vocalization "etto" (a common CF in Japanese comparable to, the authors state, the English "well..." and "uh....") Their results are worth recreating here:

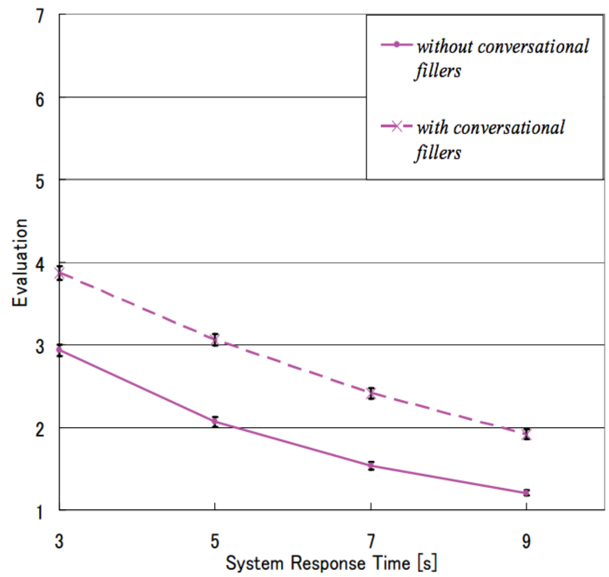


Figure 1: User evaluations across different response times and conditions. Filled pauses were preferred at every point (Kanda et al., 2009).

The results clearly indicate that CFs can be effective at raising subjective evaluations of robot response delays. It is important to note that the participants were asked only to rank their "preference about the timing of the system response" from one to seven. Not tested were any other effects that CFs had on the nature of the interaction.

CFs have also been studied as part of not robotic but 'embodied conversational agents' or ECAs. ECAs are software avatars that are designed to resemble a human and can hold conversations with them. Less promising results have come from Pfeifer et al's work on CFs and ECAs from 2009. In this study various aspects of participant attitudes towards CFs were tested, including satisfaction, trust, likability, knowledgeableness, and naturalness were compared against agents that did not use CFs. The results seem to show no clear preference for one of the other agents. The authors conclude by saying that the work is in its infancy and should act solely as an inspiration and guideline for future research. However, the study does make clear that researchers have begun being interested in what effects other than simple subjective evaluation the use of CFs has in HRI.(Pfeifer and Bickmore, 2009)

As mentioned above, CFs have also been studied for their use as aiding a WoZ when forced into actions that create delayed robot responses. In a study by Glas and Kanda in 2012, the operator's estimation of response delay duration and perceived workload were compared across conditions with automatically triggered versus manually triggered CFs. They found that automatically-triggered CFs were effective at allowing the WoZ to better estimate the elapsed delay time and that auto-fillers did not have an effect on perceived workload when compared to manually triggered CFs.

Better time estimation during (auto) filled pauses, the authors admit, is most likely due to the fact that their fillers have a fixed duration and were repeated in regular intervals. When the operator knows the approximate duration of his robot's CF she can use the CF as unit of time measure. Overall, the research motivates further investigations into the effects CFs have on operator workload and delay duration estimation. (Glas et al., 2012)

It is clear from these studies that the merits of employing CFs in order to ameliorate the negative effects of robot response delay in WoZ of scenarios are to some extent understood by the HRI community. Conversely, lacking from this research, we have seen, is a systematic study of the effects beyond mere preference when employing CFs to fill WoZ-caused interaction gaps.

2.5 HRI vs. CRI

For reasons ultimately unclear, young children tend to form a social connection more easily with robots than do adolescent and adult interactors. Children between eighteen and twenty-four months have been shown to form lasting social bonds with robots that actually improve over time (Tanaka et al., 2007). In this study, an experimenter manually assessed the quality of an interaction between a rather simple humanoid robot and the young children by simply turning a dial (a method borrowed from marketing research called "continuous audience response methods"). They were surprised to find that so long as the robot had varied, unpredictable behaviour, children of this age group had the dial turned up for prolonged periods of time and readily adopted a care-taking attitude towards the robot, treating it as a peer for which their fondness grew over time. While the authors of this study attribute these young children's eagerness to bond with the robot to their lack "preconceived notions of robots", ALIZ-E researchers contend that the situation is more complicated (Ros et al., 2011). In this paper, the authors contend that success with children of this age "may well reflect the predominance of imaginative play as a means for exploring and learning about the world in early childhood", a theme we will return to in the next section.

The situation becomes more challenging, however, when trying to establish long-term social bonds between robots and older children. For instance, Kanda et al express the difficulty they encountered with children aged ten and eleven years. They found that while some children maintained interest in their humanoid robot over their two-month experiment, two-thirds had lost interest by the experiment's conclusion, usually by week five to seven. Children seem to still be very excited by the robot at first but are much less likely to treat it as a "social actor" as they reach adolescence (Ros et al., 2011). It is clear that sustained interaction with adolescents is possible, only that special care and strategies need to be developed and followed to make long-term interaction possible.

The situation is most dire with adults. Not only has the "make believe" stage of their life usually passed, but adults typically come loaded with preconceptions of robots (popularized by film, science fiction, and the like) that tend to be quickly disappointed when they encounter today's robots (Ros et al., 2011). The result has been a rather sharp division in the types of research being done in HRI, with Child-Robot Interaction forming an almost separate discipline: CRI.

Overall, these points suggest that while convincing adults that robots ought to be approached as social entities seems outside the realm of possibility given today's limitations in AI, and that very young children tend to be, as Ros et al say, the "low-hanging fruit" of HRI research, the development of HRI techniques effective at enriching and prolonging the interaction between robots and adolescents seem to be of the highest interest. The current work supports this pursuit by investigating whether employing CFs succeeds at convincing children that a robot is more lifelike (and therefore worthy of social considerations) while at the same time introducing behavioural variety that keeps engagement high over longer periods of time.

2.6 Believability, Make-Belief, and the Intentional Strategy

Robots have been built that seek to encourage those who interact with them to believe, or at least pretend, that these robots possess a mind of a sort analogous to those we suppose animals or other humans posses. For instance, Sony's AIBO has been carefully designed to be dog-like. Apart from resembling a dog, it has been programmed to bark, whine, and play fetch. The relationship that owners of these robots have with their AIBO has been studied by analyzing the nature of postings to an AIBO online forum (Friedman et al., 2003). This study has shown that owners often speak of their gadgets as though they have mental states and a life-like essence. The work suggests that the AIBO's goal of being understood in some way similar to how we understand the nature of our pets has to a certain extent been successful. The AIBO is a clear example of the creative use of simple, non-mental, rule-based processes for a masquerade of a dog-like essence.

There is a strong chance that the chatter was, at its base, playful. The authors of the study admit that "...we are not saying AIBO owners believe literally that AIBO is alive, but rather that AIBO evokes feelings as if AIBO were alive" (Friedman et al., 2003). The attribution of a life-like essence could be a form of "make believe", that is, willfully treating the AIBO as though it had intentionality (beliefs, desires, intentions, etc.) while at once recognizing that the AIBO possessed none of these things (74% of posters had at least once referred to the AIBO as a technological artifact e.g. "a cool piece of technology").

The same issue arises in child-robot interaction experimentation. Whether or not a child truly believes that a robot has intentionality or whether they simple make believe that they do is notoriously hard to make clear. An illustrative example comes from Kahn and Friedman's work on children's differing attitudes towards an AIBO and a stuffed dog.

> The results highlight the difficulty of demarcating children's "real" from imaginary judgments...For example, while about half the children said that AIBO and the stuffed dog could hear, if they really believed it then one would have expected children to use verbal directives to AIBO and the stuffed dog in about equal proportions. However, the results showed that children used more verbal directives to AIBO (54 occurences) than to the stuffed dog (11 occurences). (Kahn Jr et al., 2004)

Believability of robots, however, need not be restricted to situations where interactors *actually* believe the robot to posses intentionality. Their believability should instead hinge on whether or not the, to borrow Daniel Dennett's famous term, intentional strategy is taken by its interactor and not whether the interactor, if pressed, believed in the actual existence of mental states in the robot (Dennett, 1989). The intentional strategy, for Dennett, is a predictive strategy whereby one takes non-realist constructs and use those constructs in predicting a thing or person's behavior. To predict that someone will be joyous because they believe they won the lottery is not to commit oneself to the idea that a 'belief' is a real, physically extended object. It is a construction of the mind that is used to explain one's behaviour while at once admitting that the construct is not a physically extended, 'real' object in the world (such as a chair or a desk). It is a predictive tool.

> ...First you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many but not all instances yield a decision about what the agent ought to do; that is what you predict the agent will do.(Dennett, 1989)

Applying this reasoning to the concept of believability we end up with a more nuanced definition. It need not be the case that a child *actually* believes that a robot is alive in possession of a 'mind'; it suffices that an interactor readily adopt and maintain the intentional strategy and have the predictions it brings forth rewarded in order for a robot to be considered believable as a possessor of agency.

A vivid illustration of this comes from work by John Harris and Ehud Sharlin (Harris and Sharlin, 2011). In this study they studied what attitudes fellow university students would have towards abstract movements performed by a preconception-free robot. They built a robot called "The Stem" which consisted of no more than a one meter long balsa wood shaft attached to a spherical joint equipped with servo motors allowing for rolls, pitch, and yaw about a single base point (picture the range of motion one would have twirling a stick, held at the extreme of one end, using nothing more than one's wrist). Participants observed a predefined set of motions (designed to be, for instance, 'happy', 'threatening', etc.) and were queried about their experience. Despite the 'robot' consisting of nothing more than a moving stick, "A large majority of partipants made at least one comment attributing an internal thought process or intentions to *The Stem*; at different times claiming the robot was "pensive... it's thinking about something" (Female, 25), "enjoying this, sort of purring like a cat" (Male, 50), hiding something (Female, 20), bowing or greeting them (6 participants)" etc. From the simple motion of a stick almost all of the (educated adult) participants adopted the intentional strategy as a predictive strategy. It is unlikely that these participants felt that the stick was *actually* thinking or greeting them autonomously; nonetheless, they (without instruction) readily adopted the intentional strategy to explain the motion they witnessed.

'The Stem' was believable as an intentional agent while, at the same time, not believed to be one.

The situation resembles the believability issues arising in work in the world of cartoon animation. Joseph Bates, an AI researcher, remarks "[AI] researchers attempting to create engaging, apparently living creatures may find important insight in the work of artists who have explored the idea of believable character" (Bates, 1994). Creating a successful cartoon character creates an illusion of agency in the beholder of an animated film. Animators do not face the issue of making their cartoon characters seem real or actual. Instead, they face the problem of making their characters believable as though they were thoughtful, living characters they pretend to be. They are tasked with making characters that are suitable for complex make-belief, that is, as the objects of an intentional strategy. The same can said to be true of robots intended to interact with humans in a social manner.

Much of the present work asks whether the use of conversational fillers enhances or diminishes an anthropomorphic robot's believability when faced with IoA-diminishing response delay. As we have seen, the success of a robot's believability can be seen as independent of whether or not the believability is simply an intentional strategy or a wholehearted ontological assertion.

2.7 Anthropomorphic Robot Movement Design

Strictly speaking, the term "conversational filler", in linguistics, refers only to the *vocalized* (or signed) utterance of a certain kind (non-lexical, time-buying, etc.). Insofar as the current work's ultimate aim is to study WoZ-aiding techniques, a looser definition will be followed. This looser, but nonetheless loyal definition, extends CFs to include iconic gestures that can aid in filling 'gaps' in HRI introduced by the adoption of a WoZ-infused interaction design.

Gestures have long been studied in HRI research as a means of conveying unspoken meaning from robot to interactor (among other uses). Nowhere is this most important than when employing anthropomorphic robots that have several degrees of movement. Gestures have been shown to be effective at drawing the interactor into a conversational interaction without disturbing the interactor's chain of thought (Tojo et al., 2000). Work by Kanda in 2007 successfully demonstrated that human-like gestures, when tastefully employed by a humanoid robot, are effective at conveying attentiveness, sympathy, and reliability (Kanda et al., 2007).

Some of the research in this field had a direct effect on the design of the CFs used in the current research. Tojo suggested in 2000 that head tilts should accompany CFs to best convey pensiveness while a lift of a humanoid robot's head of three degrees while speaking can mitigate the problems created when using a robot without a mouth (Tojo et al., 2000). Synchronized arm movements and eye contact were shown to be of partcular importance by Kanda in his 2003 work utilizing a motion tracking system (Kanda et al., 2003). The robot in this study was set to break eye contact when uttering a filler, to use coordinated movements of its arms, and to lift its head when making an utterance.

Children have been shown to be quite sensitive to movements characteristic of moving humans and animals (Fox and McDaniel, 1982). The impression is immediate and robust. Additionally, a human's (or animal's) intention has also been seen to be easily divined by children from a being's movement and can be immediately differentiated from mere random movements (Okita and Schwartz, 2006). In a particularly striking study by Meltzoff in 1995 18 month-old children observed adults (purposefully) failing, for instance, to drop beads into a cup. The infants were then observed to complete the *inferred, unseen goal* of the experimenter (as opposed to a rote reconstruction of their movement). This, and related research, strongly suggests that children are able, from a very young age, to infer a goal from human (or animal) movement. Interestingly, when the same failure of dropping beads into a cup was done by a mechanical set of pincers, the child did not seem to infer the intended goal(Meltzoff, 1995). Therefore, if children infer goals or intentionality from movements it may be seen to attest to the personification of an anthropomorphic robot.

Capitalizing on these revelations, the present work incorporated many iconic movements that often accompany either acknowledgement (e.g. nodding) or pondering (e.g. scratching the back of the head). The intention was to emphasize the intent of the accompanying vocal CF (e.g. "aha" for acknowledgement and "hmmm" for pondering), namely that the robot acknowledged receipt of information or that it is taking a moment to think its response over. (For an exhaustive list of employed movements and vocalizations see the "Conversational Filler Design" section below).

Introducing movement at strategic points in human-robot interactions can have another important benefit to the overall quality of the interaction. It has been shown that when interacting with a robot for a prolonged period of time, children go through three phases of interaction dynamic: "great excitement", "stable interaction", and "saturation" (Kanda and Ishiguro, 2005). These three phases - observed during the interactions in the present work - describe how a child is initially ecstatic to interact with a robot but that this initial excitement tapers off as the interaction progresses. The novelty of the robot quickly fades away as time goes on and is replaced by, eventually, boredom. Takayuki Kanda, a very experienced child-robot interaction researcher from Kyoto, has found success combating this effect by reserving some of the robot's behaviours for when the child becomes 'saturated', that is, when they have become so familiar with the interaction that they become disengaged. CFs incorporating movements are well suited to this aim. A careful schedule of CF deployment could keep some CFs in store until effects of saturation are observed. A discussion of the effectiveness of using CFs in this manner is presented in 3.7.

The movement of anthropomorphic robots must be carefully managed to bolster, and avoid diminishing, the illusion that the robot has human-like agency. Children are very sensitive to biological and intentional movements which provides a straightforward opportunity to increase the perceived aliveness of humanoid robots such as the Nao. As we will see, these movements can also have the added benefit of prolonging the duration of a child's excitement and engagement with the robot.

3 Conversational Filler Design

As previously mentioned, CFs were custom-made for the purposes of this experiment. As the experiment made use of a Nao, an anthropomorphic robot, particular care was taken to construct CFs that seemed human-like and lively. Each CF was a combination of a vocalized utterance and a movement. The movement was typically a combination of a moving of the head and arm(s). Two types of CFs were used: acknowledgment fillers (1.5 second fillers that signals a successful receipt of information) and pensive fillers (7-second fillers that indicate that the robot is 'thinking' of his response). These will be discussed separately.

3.1 Acknowledgment Fillers

These fillers combine some kind of robot utterance (lexical or non-lexical) with an iconic movement designed to reflect the robot's successful receipt of information and its intention to take the speaking turn. One of four iconic movements were randomly paired with either one of two non-lexical utterances or one of four lexical utterances. Lexical utterances had a fifty-percent chance of being selected. During the course of all movements the robot's eyes would flash. The pitch of its utterance would be varied slightly as would its vocalization rate in an attempt to introduce further variety.

Non-Lexical Utterances The speech synthesizer was fed a series of phonemes in place of regular word to avoid unacceptable pronunciation. The utterances were:

Phonemic Series	**Approximation**
m m m	"mmm"
a h a	"aha"

Lexical Utterances Translated into English, the four lexical utterances are as follows:

1. "Good"
2. "Ya"
3. "Nice"
4. "Very good"

Iconic movements

1. **Arms Up**
   Both arms were quickly raised and lowered to the side of the robot's body
2. **Nodding**
   A fast up and down movement of the head
3. **Fast Nod**
   A fast, single up and down movement of the head
4. **Head Up**
   The head was raise up and slightly to one side.

3.2 Pensive Fillers

These fillers combine some kind of robot utterance (lexical or non-lexical) with an iconic movement designed to reflect ongoing thought. One of eight iconic movements were randomly paired with either one of nine non-lexical utterances or one of four non-lexical utterances. Lexical utterances had a thirty-percent chance of being selected. During the course of all movements the robot's head would also move gently side-to-side and its eyes would flash. The pitch of its utterance would be varied slightly as would its rate in an attempt to introduce further variety.

Non-Lexical Utterances The speech synthesizer was fed a series of phonemes in place of regular word to avoid unacceptable pronunciation. The utterances were:

Phonemic Series	**Approximation**
a h h	"ahh"
M m m m	"Mmmmmm"
A m	"um"
Y m m	"Uummm"
h Y m m	"Hmmmmm"
h u m	"Hum"
y a	"Ya"
U m m	"Ommm"

Lexical Utterances Translated into English, the four lexical utterances are as follows:

1. "Good Question"
2. "Let me think"
3. "I thinnnnk"
4. "Now" (a common CF in dutch)

Iconic movements

1. **Hand to back of head**
   The right arm is brought to the back of the robot's head. Its open-palm hand then begins moving up and down by means of a bending at the elbow.

2. **Looking up**

   The head is angled up and to the robot's right. The hands, at the robots side, open and close and move slightly.

3. **Hand to chin**
   The robot's right hand is brought directly beneath its chin. It then opens and shuts its hand

4. **Arm to head**
   The right arm is brought with its elbow raised to the top of the robot's head. Its hand, positioned at its heads center, would then quickly open and close its fingers, pausing between cycles. This gave the impression of the robot scratching its head.

5. **"Mr Burns"**
   The robots two hands are brought to meet directly in front of the robot. The fingers of each hand would slowly open and close quickly and repeatedly.

6. **Side shrug**
   The robot's arms are made to slowly move away from its body, at which point they would quickly snap back to their original position and repeat the movement.

7. **Arms think**
   The robot's right arm is extend from its chest with its palm facing downward. The hand then is slowly rotated to face upward at which point it would quickly reset itself to face downward (as though it is weighing options). Its left arm would move as in the 'Side Shrug' described above.

8. **Hand out front**
   The robot would extend its right arm alone and move it side to side while rotating its hand at the wrist.

4 Experiment

4.1 Introduction

The experiment sought to 1) confirm that CF-filled robot response delays of greater than one second were preferred to unfilled pauses 2) explore what subsidiary effects CFs had on the (illusory) perception of the nature and abilities of the robot by child interactors, and 3) to measure what, if any, effects CF use had on WoZ stress or response delay duration estimation. Of particular interest was whether the use of CFs were able to improve upon (or slow the decay of) a child's perception that the robot responds quickly, is intelligent, shows self-directed behaviour, is likable, and seems more alive. Furthermore, it was asked whether these advantages could be attained without the added cost of making the robot seem less intelligent, trustworthy, friendly, amicable, or human-like in the eyes of child interactors.

The experiment divided twenty-six children into two balanced groups (see 'Participants' below for a detailed account) that each interacted with a Nao robot [see the appendix for an in-depth description] in two sessions. The two sessions were near-identical in format to each other, different only in the presence or absence of conversational fillers (see 'Experiment Design' for details) and distinct identities for the physically identical robot actors. Each child encountered two robot characters (one for each session): 'Charlie' in the first session and 'Robin' in the second. In fact, the same robot was used for each session; the differentiation was made only by changing the robot's shirt and with slight differences in the introductory portion of the opening dialog (see 'Experiment Design' for more detail).

While all fillers served to 'buy time' for experimenter action, they came in two sorts. The first, used in the introductory phase of a CF-filled session, consisted of a non-lexical 'grunt' of an affirmative nature and, sometimes, 'okay' , e.g. "ah, okay", heretofore referred to as an 'acknowledgement filler'. This short filler (One-and-a-half-seconds) was accompanied by a rapid nodding of the Nao's head. The second variety, referred to as a "pensive filler" was a single linguistic or non-linguistic utterance of a pensive sort accompanied by one of seven different iconic thinking motions (e.g. scratching of the head or chin). This filler, during the dialog portion of each session, was seven seconds in length.

Gleaning useful measurements from children is notoriously difficult, due largely to their desire to please the experimenter by providing what they perceive to be desired evaluations, referred to as "demand characteristics" (Orne, 1962). Previous attempts within the ALIZ-E project have seen this manifested with children, for instance, selecting responses on questionnaires solely in the extreme positive (Tielman et al., 2014b). In light of these issues several precautions were taken in the design of questionnaires administered to the children. In the questionnaire given at the end of each session, the "Post-Trial Questionnaire" (see the 'Measures' subsection below for a detailed account) was specifically designed to combat this effect. In previous ALIZ-E experiments a five-point scale was used for each participant response, with each selection accompanied by a 'smiley face' (caricature drawings of human faces varying only in the arch of the mouth) ranging from a frown, neutral, to a large smile. This method was abandoned in favour of a simple three-point evaluation for most questions with no accompanying graphic; simply "nee" ('No'), "Misschien" ('Maybe'), and "Ja" ('Yes'). The choice, while not necessarily supported by research, was made to disassociate feelings of negativity linked to providing a non-extreme response in the positive. It was hypothesized that the addition of 'smilies' inadvertently amplified the effect of demand characteristics. It was stipulated that actual sentiments of less-than-extreme nature were masked by a child's desire not to indicate anything but total enthusiasm to the experimenter. Using a three-point scale as opposed to the five-point scale used previously in ALIZ-E experiments also amplified the measured effect of non-extreme responses (participants that chose a four-out-of-five level response is statistically less powerful than a choice of two-out-of-three). This choice of questionnaire design, as we will see, seems to have been successful in mitigating the effect of demand characteristics.

In addition to this consideration, the "Final Questionnaire" (see 'Measures' for a complete description) consisted entirely of a forced-choice between 'Charlie' and 'Robin', the robots used for the first and second session, respectively, for each participant. While this method does not guarantee the absence of demand characteristics (the child may have ascertained that the robot employing CFs was the robot the experimenter 'wanted' the child to choose) the strategy seems to have been successful at minimizing this effect. This is seen by the fact that children's responses to many of the forced-choice questions followed a near-random fifty-fifty distribution (see 'Results' for more details).

Most notable in this respect, however, was the use of covert means of implicitly measuring a child's liking or preference of/for one of the robots. The 'Free Play Paradigm', used previously in ALIZ-E experiments with children, has the experimenter leave the room, under false but benign pretenses, leaving the child with the option of continuing to engage the robot (in this case continuing the quiz with the robot) or some other entertaining options (see 'Design' for more detail). In addition to this, at the end of the second and final session with each child the child was offered the opportunity to have their picture taken with one, and only one, of the two robots. Each of these methods, while not having concrete support from prior research, were assumed to be implicit measures of the likability of the quiz interaction, in the case of the 'free play' measure, and the child's overall preference for one of the two robot characters for their choice of photo partner.

Also on the Final Questionnaire was the option for the child to explain their (forced) selection of one of the robot characters (prompted by a simple 'waarom?' ['why?'] after each of the six questions. See 'Measures' for an exhaustive list). This addition was prompted by the casual but insistent recommendation given by a previous ALIZ-E researcher emphasizing the value of open "why?" questions in understanding children's responses. Highlights from these responses are presented in the 'Results' section below.

As a whole, the measures on the Post-Trial and Final Questionnaires, along with the covert measures discussed above, sought to explore both preference for CF-filled (as opposed to unfilled) robot response delay and the subtleties that accompany these two conditions. The measures were designed to ascertain the difference in perceived:

1. **Responsiveness**
   The perceived speediness in robot response independent of its actual response speed. While the filled gaps do not shorten the actual duration of response delays in this work, we hope to see an increase in the child's sense of speediness when CFs are employed.

2. **Likability**
   The child's overall sentiment towards the robot, ranging from extremely positive affect to extremely negative.

3. **Aliveness/Humanness**
   Children, as discussed at length in the 'Background' section of this document, tend to overgeneralize simplistic ideas of biology and aliveness from the animal world to (wrongfully) include robots.

4. **Agency / Self-Directed Nature** Agency refers to beings that weigh possible actions to achieve self-conceived goals. The choice to involve a WoZ in influencing robot behaviour tends to reflect the experimenter's desire to inflate the impression of a robot's agency (as using a WoZ tends to indicate a desire for robot behaviour or intelligence that is currently technologically impossible. This is by no means a rule, however).

5. **Intelligence / Competence** As robots are often used to educate the participants, as in the ALIZ-E project, an increase in perceived robot intelligence on the part of the participant may have important consequences on learning and trust.

6. **Trustworthiness** This could be rephrased as the perceived likelihood that the robot would be moved to deceit, that is, to lie. This is to be conceptually separated from the trust the participant feels due to the robot's perceived expertise (this is covered by 'Intelligence / Competence' above).

It is important to stress that the work sought not only to measure *improvements* in these domains but also any accompanying *deterioration* of these aspects of a child's perception of a robot employing CFs.

4.2 Experiment Design

Two groups were used for a two-by-two, balanced, within-subject experimental design. Each participant took part in two interaction sessions with a Nao anthropomorphic robot, each corresponding to a conversational-filler enriched condition (called the "CF" condition) and one devoid of conversational fillers (called the "Non-CF condition"). Each interaction session followed an identical format. Each session consisted of two components. The first was a structured, pre-scripted dialog with the robot (named 'Charlie' for the first session for each group and 'Robin' for the second, discussed in the 'Introduction' above).

The dialog began after the child was seated approximately a meter from the robot. When the experimenter triggered the beginning of the session, the robot would stand and become animated by neutral movements [see the appendix for an in-depth description]. Using the speech synthesizer and in-built speakers of the robot, the robot would introduce his/herself (the gender was purposely left obscure) and ask the child's name. After the child provided their name, the experimenter entered the information into the user model software and, when finished, triggered the next portion of the dialog. The child's age and hobby were ascertained and recorded in a similar fashion. In the CF condition each of the participant's responses were followed by a non-lexical acknowledgement filler; the Non-CF condition had the robot continue its neutral movements during the time it took the experimenter to enter the information into the user model, approximately four seconds. The introductory portion of the dialog was followed by the robot prompting the participant to ask it a question (of the participant's design) three times. When the child finished asking the question the experimenter would prepare a custom response. The response delay had an enforced minimum duration of seven seconds and very rarely extended beyond this minimum. If the child refrained from asking the first of the prompted questions, the robot would verbally encourage the child to do so; the child was allowed to abstain from posing the second and third question without encouragement from the robot.

The second portion of all sessions consisted of the robot and child playing a quiz game together with the aid of a shared tablet computer sitting between them. The child and robot took turns asking the other questions displayed on the tablet computer [see appendix for complete list of questions]. The tablet was affixed to a 'seesaw', that is, was held semi-upright by an angled platform that could be reversed with a push.

Twelve quiz questions were posed in total, alternating between participant and robot as the actor asking the question. The robot always took the first turn and knelt and extended its arm to flip the table to the child when the change from robot to child as question asker was needed. If the child did not turn the tablet before the time came for the robot to pose a question to the child the experimenter triggered a vocalization asking for the tablet to be flipped. This helped maintain the illusion that the robot, like the child, was reading the question from the tablet (in fact it was not).

After the child read the question from the tablet aloud for the robot, along with the four multiple-choice possibilities, the robot (triggered by the experimenter) would, in the CF condition, perform a conversational filler. In the non-CF condition the robot would perform neutral movements for the identical amount of time as the CF was performed in the CF session. The robot then provided its answer. The probability of a correct response on its first attempt was sixty percent across all questions and conditions. Should that response not be correct the robot would randomly choose from the remaining possible answers, continuing until the correct response was produced or the child cut the robot off.

On the six occasions in each session where the child posed the robot a question and the robot had its turn to respond, a delay always proceeded the robot's response. The duration of the delay varied for each question (but remained the same for each repeated attempt at answering within a question). The delay had a mean of five seconds and a standard deviation of one second [see the appendix for a complete list of delay durations]. This set of response delays was fixed for each session for each condition. The times were chosen to be in accordance with the suggested maximum acceptable delay suggested by Shiwa et al. (seven-seconds) in their seminal work on CFs and human-robot interaction discussed elsewhere (Shiwa et al., 2008) while providing some variety.

It is important to emphasize that regardless of whether the robot was responding in a CF or Non-CF condition the duration of the delay between the end of the child asking the question (or informing the robot of its incorrect response) and the beginning of the robot's spoken answer was fixed for each question. Restated, regardless of whether or not the robot employed CFs in the session there always existed a delay of predetermined length before providing the robot's response to the participant's question.

After all twelve questions were asked the robot informed the child that the quiz was complete and expressed his hope that the child had a good time. The robot then knelt down and became inanimate. The experimenter then approached the child and concluded the interaction. At this point a 'free choice' period was offered to the child under the false pretense that the experimenter needed to leave the room to perform a duty. The child was informed that he or she had the choice of continuing to play the quiz with the robot, to read some nearby comics, or to play Sudoku. Their choice in activity was recorded and used as a measure of implicit enjoyment of the quiz activity (discussed in the 'Measures' section).

Each trial concluded with the administration of the 'Post-Trial Questionnaire' (discussed in 'Measures' below) on the reappearance of the experimenter from their absence from the room to accommodate the free play portion of the experiment. The second and final session then saw the administration of the "Final Questionnaire" (also discussed in 'Methods'), followed by an aliveness ordering task in which the children ranked a number of inanimate and animate objects, along with the two robots, from least to most alive (see 'Methods'). Finally the children had an opportunity for the child to pose for a photograph with one-and-only-one of the two robots. The choice of photo partner was taken as a measurement of implicit preference for one of the two robots.

Separate from the work discussed above, the children conducted two related tests: the Pre- and Post-Experiment Knowledge Tests. Half of the questions were drawn directly from the quiz and were identical on each of the two tests. The other twelve questions were unrelated to the rest of the experiment and functioned only to obscure the importance of the other twelve (see 'Measures' for a more detailed explanation). The first test was administered a week prior to the beginning of the experiment and the second approximately a week after the experiment's conclusion. The difference in individual children's scores were calculated and it was tested whether children were more likely to learn the answer to a quiz question if that question appeared in the child's CF session.

4.2.1 Participants

Twenty-six nine to eleven year-olds (n = 26, avg = 9.32 years-old) from a single class of Dutch sixth graders from the Postiljon school in Soesterberg in the Netherlands were used as unpaid participants for the experiment. They were evenly divided between two groups balanced by age, gender, and self-reported liking of computers, the last being a balance suggested by researchers in a related experiment (Pfeifer and Bickmore, 2009). The children were introduced to the robot approximately a week prior to the experiment during a brief presentation by the experimenters. Excitement was high as the children had been shown pictures of the Nao robot [a detailed description of the Nao robot can be found in the appendix] and informed of their chance to do an activity with the robot. This is worth noting as prior expectations, as we will see in detail in the 'Discussion' section to follow, can have important ramifications for the experience children have when interacting with robots. Research has shown that high expectations, informed by fiction's depiction of robots and the robots' anthropomorphic appearance, are often met with marked disappointment when robots fail to meet these expectations (Komatsu et al., 2012). This effect seems to be responsible for a not insignificant ordering effect observed in the current research examined at length in the 'Discussion' section to follow.

4.3 Methodology

4.3.1 Nao Robot

The Nao robot was the sole robot employed for the experiment. The Nao is a small humanoid robot 58cm in height with 25 degrees of freedom (e.g. movable head and limbs) weighing 4.3kg. It is equipped with a wide array of sensors and actuators: 4 microphones, 2 cameras, a gyroscope, an accelerometer, 2 speakers, and 2 IR and sonar sensors. The Nao is equipped with WiFi and Ethernet and runs a custom built form of Linux called "Naoqi".

4.3.2 Experimental Flow

This section acts to disambiguate the details of the experiment introduced in the 'Experiment Design' section above and to clarify the role of the WoZ. The session begins with the child sitting approximately a meter from the inactive Nao robot crouched on a table. The session began when the WoZ loads the WoOz software on a laptop at a desk behind, and out of sight of, the child participant. From this moment until the end of the session the Nao performed neutral movements (apart from when other movement commands overrode these movements), that is, small movements of its head, arms, and legs. These movements were inherited as part of the ALIZ-E software package and serve to animate the robot. The express purpose of these movements is to give an impression of aliveness or animacy. The session, mentioned in the 'Design' section above, was divided into two sections: the dialog and quiz sections. First came the dialog section.

Dialog Section The purpose of this section of the interaction was to employ CFs in their destined role as a WoZ aid in conversation. The dialog came in two parts. Firstly, an introductory portion in which the Nao introduced itself and queried the child for three pieces of information: their name, age, and hobby. After the child's response for each question the experimenter hit the [F1] key, triggering a 2-second CF in the CF condition and enforcing an unfilled pause of a minimum of 2-seconds (see 'Software Design and Contributions' for a description of this mechanism). For example, the first session began as follows (translated to English here):

[Robot stands, neutral movements begin]

Robot "Hello, my name is Charlie. I am here to play a quiz with you. You will really enjoy it. Before we begin I'd like to know more about you. First of all, what's your name?"

Participant "John"

WoZ [F1] tapped. One-and-a-half-second minimum pause started. Acknowledgment CF activated in CF condition. 'John' entered in User Model Window (see 'Software Design and Contributions' for more details on this window)

Robot "John. I find that a very pretty name. How old are you, John?"

Participant "I am eight years old"

WoZ [F1] tapped. One-and-a-half-second minimum pause started. Acknowledgment CF activated in CF condition. Eight entered as age in User Model Window.

Robot "Ah you are already eight years old! What do you like to do after school?"

Child "Playing Ping Pong"

WoZ [F1] tapped. One-and-a-half-second minimum pause started. Acknowledgment CF activated in CF condition. 'Ping Pong' entered in User Model Window.

Robot "I also find playing Ping Pong nice"

At this stage the dialog continues seamlessly to the second part of the dialog begins. In this section the child is asked if he would like to ask the robot any questions. If the child did not volunteer a single question a single verbal encouragement was offered. For instance:

Robot ”What would you like to know about me?”

Participant ”I don’t know...”

Robot ”You can’t think of anything?”

Participant ”What is your favourite sport?”

WoZ [F9] tapped. Seven-second minimum pause started. Pensive CF activated in CF condition. During this time the WoZ types a custom response on the dialog window (see ’Software Design and Contributions’ for more detail). After a minimum of five seconds the WoZ tapped the [send] button and the response was vocalized by the Nao as follows.

Robot ”I think football is my favourite sport. What else would you like to know?”

Child ”Do you also like hockey?”

WoZ [F9] tapped. Seven-second minimum pause started. Pensive CF activated in CF condition. During this time the WoZ types a custom response on the dialog window and sends it after at least five seconds.

Robot ”Yes, it is also great. Is there anything else you would like to ask me?”

Child ”No.”

Robot ”Okay. We are now going to play a quiz together. We are going to take turns asking questions off the screen. The other person has to give their answer out loud. I will go first. John, are you ready to start the quiz?”

Child ”Yes”

Thus concluded the dialog portion of the interaction. In it was a total of 6 pauses (with or without a CF), three one-and-a-half-second pauses and a minimum of one and a maximum of three five-second pauses (with or without a CF).

Quiz Section Directly following the conclusion of the dialog section of the interaction the robot and child would begin playing the quiz game together. The same twelve questions were used for the first session for both conditions, likewise for the second session. Each question the child posed the robot had a predefined delay before the robot responded. The pause was filled with a pensive CF in the CF condition and had the robot sit idle, apart from neutral movements, during the Non-CF condition. The duration of that pause was defined beforehand for each question. The set of pause durations was the same for each session. A typical interaction, translated here into English, was similar to the following:

Robot What is the capital of Canada? Is it A, Toronto. B, Ottawa. C, Montreal. Or D, Vancouver.

Child "I think A, Toronto"

WoZ A entered into Quiz Window as participant response.

Robot "John, I am afraid this is incorrect. Please try again."

Child "D, Vancouver?"

WoZ 'D' entered into Quiz Window as participant response.

Robot "This is also not correct. The Correct response was B, Ottawa". The robot would then kneel and flip the tablet to face the participant.

Child "What is Beyoncee's last name. Is it A, 'Lopez'. B, 'Knowles'. C, 'Fierce'. Or D, 'Gaga'?

WoZ [Give robot response] button clicked. A pause of a predestined amount of time followed, either filled with a CF in the CF condition or empty (apart from neutral movements) in the Non-CF condition.

Robot "I think A, 'Lopez'"

Child ”No, try again”

WoZ [Give robot response] button clicked. A pause of a predestined amount of time followed, either filled with a CF in the CF condition or empty (apart from neutral movements) in the Non-CF condition.

Robot ”Is it B, 'Knowles'?

Child ”Yes!”. The Child turns the tablet to face the robot.

This process would continue six more times with novel questions. At the completion of the quiz the robot announces that the session is over and expresses his hope that the child had a good time.

4.3.3 Balancing Robot Response Delay

Particular attention was paid to the experiment's design to ensure that each child was exposed to the same amount of robot response delay in each session (and therefore each condition). The delays preceding the robot's response to each quiz question were fixed and identical for each session. Additionally, a minimum of one and a maximum of three child-posed questions during the dialog section of the experiment was enforced (and, as evidenced in ”Questions Asked in Dialog” in the results section below, this did indeed result in approximately the same number of questions being asked in each condition).

4.4 Implementation

4.4.1 ALIZ-E' ”WoOz” Wizard of Oz Software

Overview The experiment was conducted entirely through the ALIZ-E ”WoOz” software, a custom built software package to facilitate the plethora of activities within ALIZ-E. The software was the single interface through which the WoZ performed her duties. This included vocalization, quiz, and CF control. The Windows-based software contains separate windows used for the two sections of the experiment: dialog and quiz. In support of these two sections there was a third window containing the ”User Model”, the portion of the software into which the WoZ would enter details obtained from the child through the course of the dialog section of the experiment. Using a mouse and a series of check boxes on the main window of the software the WoZ could show or hide any of the other aforementioned windows. The modifications to these windows and their operation to accommodate CFs is discussed in the following section "CF Integration".

The Dialog Window This window was the sole window used for the dialog section of the experiment (see 'Design' above). The window is divided into two sections.

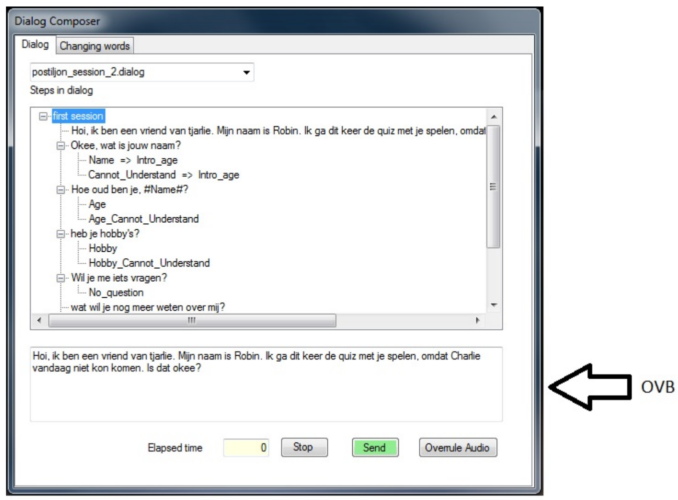


Figure 2: The Dialog Window with On-Deck Vocalization Box indicated

The bottom of the window is a large text field into which text can be placed for immediate vocalization by the Nao (and its Dutch speech synthesizer). This will later be referred to as the "on-deck vocalization box" or OVB. Custom responses (at any point in the experiment) could by typed and sent from this box with a tap of the [enter] key or a click of the [send] button. Above this box stands a hierarchically-arranged script loaded at the launch of the WoOz software. The script, discussed in the 'Methodology' section above, allowed for a controlled conversation between the robot and child. Clicking a line of text automatically filled in the on-deck vocalization box with the content of that line and was sent to the robot in the aforementioned fashion.

The User Model Window Here the WoZ could enter in certain user information ascertained during the course of the dialog phase of the experiment including the child's name, age, and hobby.

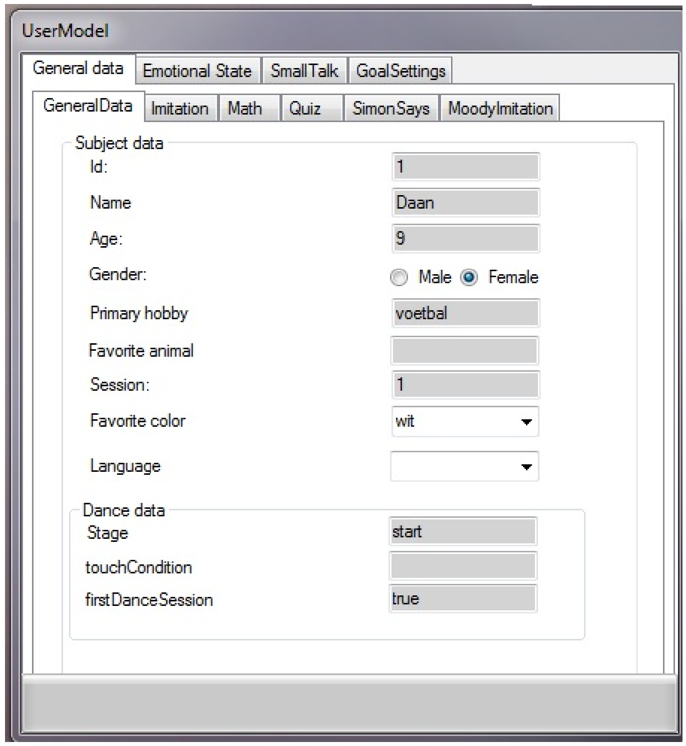


Figure 3: The User Model Window

The arrangement, it should be noted, is somewhat clumsy: it requires that the WoZ first ask the child for said information through the Dialog Window, then change active windows to the User Model Window, enter that information, then switch back to the Dialog Window to continue the conversation. As we will see in the following section on CF Integration, acknowledgement CFs (see 'Design') were used in an attempt to ameliorate the inevitable delay in interaction caused by this procedure.

The Quiz Window This window was used to conduct the quiz in all sessions. It allowed the WoZ to begin the quiz, enter the child's response to a question (A,B,C, or D), instruct the robot to ask a question, send pre-set vocalization commands to the robot (e.g. "Turn the tablet, please", "It is your turn to ask the question", etc.), send the robot's response (according to the probability of a correct response being sent discussed in 'Design' above), and go on to the next question.

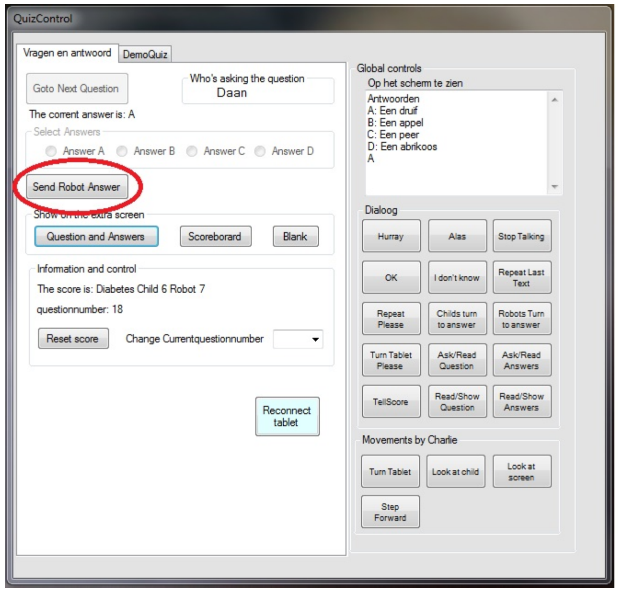


Figure 4: The Quiz Window with the [Send Robot Answer] button encircled.

The progression through the quiz was relatively painless.

4.4.2 CF Integration

Several steps were taken to fold CFs into the existing software in places that would make the experiment both possible and user-friendly for the WoZ. The experiment called for CFs to be either automatically or manually triggered alongside existing functionality. Acknowledgment and pensive fillers had to be easily activated at user-controlled but predestined moments during the dialog phase of the experiment. For the quiz phase, pensive fillers were required immediately following the completion of the participant's provided answers to quiz questions.

Furthermore, particular attention was paid to ensure that identical system response delays were present for paired questions regardless of whether a CF was performed or not (to make differing impressions of the robots' (identical) response speeds interesting) during the quiz. The same was required for the dialog phase (i.e. regardless of how long it took the WoZ to produce a response in the OVB the system must have the same delay to ensure experimental integrity).

On-Deck Vocalization Box As the WoZ was the sole mechanism by which the end of a child's speech phases were determined, it was important that the WoZ have complete control of the initiation of CFs during dialog. In addition to the design constraint of easy ignition of CFs, it was essential that the WoZ be able to easily and quickly choose between acknowledgement CFs and pensive CFs to suit the experimental situation.

To meet these design constraints it was decided to have acknowledgement fillers triggered by a tap of the [F1] key and pensive CFs triggered by the [F9] key. These keys, on standard keyboards, are the first of four in the two groups of function keys.

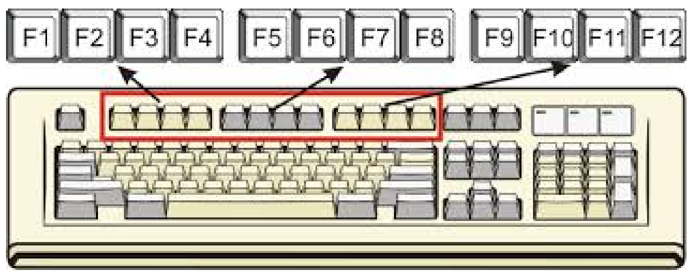


Figure 5: Function Keys From Prototypical Computer Keyboard

This permitted quick and easy search for the keys while keeping them both physically and conceptually (as members of different groups of keys) separate. Since acknowledgement fillers are shorter than pensive ones it also permitted the mnemonic device of "shortest to longest" from left to right (or smaller to bigger function number i.e. 1 versus 5).

Tapping these keys while the OVB had focus had a second feature. Since pensive fillers in the context of human-robot conversation are generally used to fill the time it takes the WoZ to construct a custom response to an unexpected inquiry from its interactor, activating a pensive filler also highlighted whatever text was currently present in the OVB.

Figure 6: Highlighted OVB content from CF initialization

This permitted the immediate start of WoZ input of a hand-crafted response for the robot since any keypress while the contents of the OVB were selected would clear the contents and replace it with the whatever key was pressed. This highlighting method is to be preferred since it did not immediately delete the content of the OVB (in case that content was still wanted) but does allow for its deletion with no extra input from the WoZ.

As discussed above, steps needed to be taken to ensure that the system response delays were as similar as possible across conditions. To address this, the tapping of [F1] or [F9] deactivated the [send] button associated with the OVB for a fixed amount of time. In the case of [F1], the resultant deactivation period was 1.5-seconds (the time of acknowledgement fills). For the [F9] button the [send] button was deactivated for seven seconds, the duration of pensive fills. These deactivation times happened regardless of whether a CF was performed or not.

User Model Data Entry As discussed in the previous section, the design of the WoOz software caused a rather prolonged delay during the entry of participants' personal information. Tapping of the [F1] key when the Dialog window had focus allowed for the easy manual triggering of acknowledgement fillers for this situation. Choosing to use the keyboard allowed a division of tasks to speed the process as much as possible: with one hand on the keyboard the WoZ could easily trigger an acknowledgment CF while simultaneously preparing to switch windows with the mouse in the other hand.

Quiz As the quiz made use only of pensive CFs the restriction of easy selection of CFs was not applicable at this stage. However, it was again true that the WoZ was required to indicate the completion of a child's conversational turn, namely, their response to the robot-posed question. As we saw in the 'Methodology' section above, the WoZ was required by the existing WoOz software to click the [send robot response] button on the Quiz window.

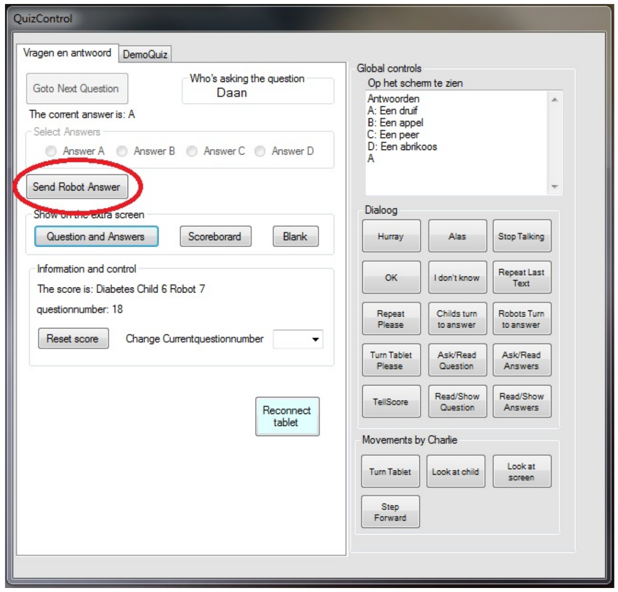


Figure 7: The Quiz Window with the [Send Robot Answer] button encircled.

It was decided to simply have the CF (or unfilled delay) triggered by the clicking of this button. This resulted in the CF or the appropriate unfilled delay being properly placed during the quiz (before providing the robot's response) without additional effort required on the WoZ's part. Once the appropriate CF or unfilled delay terminated the robot would provide its response as usual. Should the first response be incorrect, as was true roughly forty percent of the time, the WoZ would wait for the end of the participant's response (e.g. "no, that is not correct. Try again!") and would simply click the [send robot response] button again, repeating the procedure again.

4.5 Measures

The present experiment had a total of six points of measurement, each of which will be discussed at length in this section. The measurements, in chronological order of their administration, were:

- Pre-/Post-Experiment Knowledge Test
- Post-Trial Questionnaire
- Final Questionnaire
- Free Play Selection

- Aliveness Ordering Task
- Photo Partner Preference
- Heart Rate (HR) and Heart Rate Variability (HRV) Data - Child
- Two WoZ Stress Measurements including HR and HRV

4.5.1 Pre-/Post-Experiment Knowledge Test

The primary goal of the ALIZ-E project as a whole is to use a Nao robot and supporting tools and activities to educate children on how to cope with diabetes. The quiz game, used in this experiment (see 'Design' above), was conceived as an engaging way to impart knowledge on children. In line with this is the question of whether CFs detracted, enhanced, or had no perceptible effect on this goal. To this end the children were asked to answer a total of twenty-four written questions a week prior to the beginning of the experiment. The test was composed of six questions drawn from each of the two quiz sessions they would later participate in accompanied by twelve distractor questions, that is, questions not used in the experiment appearing only to mask the intention of the exam. A corresponding Post-Experiment Knowledge Test was administered in the week following the quiz with a similar design: the same twelve questions, phrased identically, used in the experiment accompanied by twelve (new) distractor questions. Each question was accompanied by a space clearly intended as the space in which the child was intended to provide their written response. There were no multiple-choice responses provided.

4.5.2 Post-Trial Questionnaire

At the completion of the quiz activity portion of each trial the participant was asked to fill in a questionnaire pertaining to the robot with which they had just interacted. The questions were of two sorts: twelve multiple choice "nee/misschien/ja" ("No/Maybe/Yes") questions and a six (unnumbered) 'number-lines'.

Multiple-Choice Questions Twelve multiple-choice questions probed the child's perception of each robot. "[robot]" was replaced by the relevant robot's name for each trial. The questions were, in order and accompanied by their English translation and the effect they were intended to measure:

1. "Ik wil graag vriendjes zijn met [robot]"
   I would like to be friends with [robot]
   -Likability
2. "[robot] lijkt me super slim"
   I find [robot] super smart
   -Intelligence
3. "[robot] maakt plannen voor het weekend"
   [robot] makes plans for the weekend
   -Agency
4. "Ik denk dat [robot] een geheim kan bewaren"
   I think [robot] can keep a secret
   -Trustfulness
5. "Ik wil graag nog een tweede quiz met [robot] spelen"
   I would like to play another quiz with [robot]
   -Likability
6. "[robot] zou een goede studievriend zijn"
   [robot] would make a good study partner
   -Intelligence
7. "[robot] gaat op vakantie"
   [robot] goes on vacation
   -Agency
8. "[robot] heeft snel op mijn vragen geantwoord"
   [robot] answered my questions quickly
   -Responsiveness
9. "Ik vertrouw [robot]"
   I trust [robot]
   -Trustfulness
10. "Ik denk dat [robot] mijn huiswerk zou kunnen maken"
    I think [robot] could do my homework
    -Intelligence
11. "[robot] gaat boodschappen doen"
    [robot] goes grocery shopping
    -Agency

12. "Ik denk dat [robot] zou liegen tegen mij"
    I think [robot] would lie to me
    -Trustfulness

'Number Lines' Four zero-to-twelve unnumbered 'Number lines' (for lack of a better term) were included in the Post-Trial Questionnaire. The ends of the line segments were labeled "Minder" ("Less") on the left side and "Meer" ("More") on the right. Number lines were used alongside multiple choice questions to measure subtle perceived differences between the two robots that would not necessarily result in a shift from, for instance, 'No' to 'Maybe' in multiple choice questions but may result in a shift when the measurement was more fine grained. For instance:

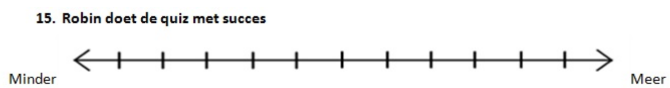


Figure 8: A sample number line from the Post-Trial Questionnaire (translation: Robin does the quiz with success)

As above, the questions are listed here along with their English translation and the character of the participant's IoA they are intended to measure:

1. "[robot] is een trage denker"
   [robot] is a slow thinker
   -Intelligence

2. "[robot] heeft snel op mijn vragen geantvoord"
   [robot] answered my questions quickly
   -Responsiveness

3. "[robot] doet de quiz met succes"
   [robot] performed the quiz with success
   -Competence

4. "[robot] was een beetje saai"
   [robot] was a bit boring
   -Likability

4.5.3 Final Questionnaire

The Final Questionnaire was administered at the end of the second and last session for each child. It consists of six forced-choice questions to which the child must choose one and only one of the two robot characters they encountered. Each choice was accompanied by a picture of the relevant robot character. Each question also provided space for the child to explain their choice, prompted simply by "Waaroom?" ("Why?"). For instance:

Figure 9: A sample question from the Final Questionnaire (translation: Which robot do you find more human-like?)

The questions are presented here in order along with their English translation and the subtlety they are meant to measure:

1. "Welke robot is beter in de quiz?"
   Which robot is better at the quiz?"
   -Competence

2. "Welke robot vind je het meest menselijk?"
   Which robot did you find most human-like?
   -Humanness

3. "Welke robot antwoordt het snelst?"
   Which robot answered most quickly?
   -Responsiveness

4. "Welke robot vind je het leukst?"
   Which robot did you find best ('leukst' can also be translated as 'nicest', 'coolest', 'most enjoyable', among others)

5. "Welke robot is slimmer?"
   Which robot is smarter?
   -Intelligence

6. "Welke robot luisterde beter terwijl jij aan het praten was?"
   Which robot listened better when I was talking?
   -Responsiveness

4.5.4 Free Play

Directly following the completion of the quiz component of each trial (in both conditions) the experimenter approached the child and informed them that the quiz had concluded. They then, in the form of a white lie, informed the child that they had to leave the room to print a questionnaire. Child was casually given the option of continuing to play the quiz with the robot, read some nearby comic books, or play the provided game (Sudoku). They were told that if they wanted to continue playing with the robot they should verbally express the desire to the robot. The options were phrased identically on each occasion and the experimenter left for exactly five minutes. If the child chose to continue playing they were presented with novel questions. CFs were employed in the CF condition in exactly the same manner and duration as the other quiz questions used in that session.

4.5.5 Aliveness Ordering Task

To assess whether and to what extent they children perceived the robots to be alive a new metric was conceived and administered. The children were presented with six randomly ordered laminated printouts of eight items after the second and last session with the robot (after the Final Questionnaire). They were then asked to order these items (by strategically laying them on the table in front of them) from "least to most alive". No further guidelines were provided nor, interestingly, were any requested by the participants. The items, in no particular order, were:

- A television
- A fish
- A houseplant
- Ninja (a popular Dutch children's book character; an anthropomorphic baby rabbit)
- A dog
- A human male
- Robin
- Charlie

4.5.6 Photo Partner Preference

The children were asked, at the end of their second and final trial, with which robot they would like to have their picture taken (if any; no participant declined the offer, however). They were only permitted to choose one of the two robot characters. While the reason for their choice was not probed, it was assumed that their selection reflected a general preference for one of the two robots.

4.5.7 Heart Rate and Variability Data - Child

For the duration of the each trial the child's heart rate and heart rate variability were measured with a wristwatch-style monitor as a possible indication of arousal and valence. The results were broken down by experimental section (dialog, quiz), averaged, then compared against a baseline measurement taken during a three-minute relaxation video (REFERENCE TO VIDEO).

4.5.8 WoZ Stress Measures

Apart from the effects CFs may have on a child's perception of anthropomorphic robots is the potential beneficiary effects their employment may have on the stress felt by the WoZ during the dialog section of the experiment in which the CFs were meant to 'buy time' for the WoZ to construct a custom piece of dialog for the robot. It is important to remember that in the experiment discussed presently the WoZ was aware that in both CF and NonCF conditions there was an enforced minimum response time for their response to children's questions e.g. regardless of whether CFs were being used to fill the gaps in which the WoZ typed responses to children's questions in the dialog section of the quiz, the WoZ knew that the gap would be (at least) the same length. This would intuitively seem to relieve the WoZ of culpability for unwanted response delay and would likely reduce stress levels for unfilled pauses. The WoZ stress was measured in two ways. Firstly, unbeknownst to the experimenter at the time, the length of typed responses during the dialog section of the experiment were compared across conditions. The assumption is that should the WoZ feel less time pressure they would type longer responses (despite being aware of the enforced minimum response time in both conditions).

The second measure was heart rate (HR) and heart rate variability (HRV). Increased heart rate is indicative of increase stress (REFERENCE) while a decrease in heart rate variability is a robust symptom of heightened time pressure situations. HR and HRV were compared across conditions. Due to experiment constraints, no relaxation exercise was undertaken by the WoZ and, therefore, no trial-by-trial baseline measurement was ascertained. While this weakens the conclusions that can be drawn from such data, it is still useful to compare averages since there were a high number of trials.

4.5.9 Demand Characteristics

While any direct querying by a researcher is a candidate for the unfortunate effect of demand characteristics (participants responding in a manner they feel will please the experimenter), covert measurements do not fall victim to this effect. For this reason two covert measures of robot preference, 'Free Play' and 'Photo Partner Preference', were used. As we will see the latter indicated an implicit preference for the CF robot while the former showed no discernible pattern.

4.6 Results

The experiment clearly suggests that the participants perceived the CF robot to be more responsive, more human-like, more alive, and more likable without the children perceiving the robot to be less intelligent or self-directed in nature (agency). The results will be presented by measurement.

4.6.1 Pre-/Post-Experiment Knowledge Test

Six questions from each quiz session were tested pre- and post-experiment. Improvements - that is, incorrect response on pre-experiment test and correct response on post-experiment test - were tallied and compared across conditions.

CF Improvements	45
Non-CF Improvements	38

While the results suggest that children were more likely to answer a previously incorrectly answered question correctly when that question appeared in the CF session of the quiz, there is a thirty-seven percent chance that the observed difference could have been simply due to chance.

4.6.2 Post-Trial Questionnaire

Sixteen written responses to twelve multiple choice and four zero-to-twelve 'number line' questions were analyzed using two two-sample nonparametric tests.

I would like to be friends with the robot	N/A	I trust the robot	N/A
Robot is super smart	.219	Robot could do my homework	.754
Robot makes plans for the weekend	.063	Robot goes grocery shopping	.125
Robot can keep a secret	N/A	Robot would lie to me	.500
I would like to play a second quiz with the robot	.500	Robot is a slow thinker (scale)	.664
Robot would make a good study partner	1.000	Robot answered my questions quickly (scale)	.134
Robot goes on vacation	.375	Robot is successful at the quiz (scale)	.077
Robot answered my questions quickly	.012	Robot was a bit boring (scale)	.180

- Results for Sign test

I would like to be friends with the robot	.317	I trust the robot	.317
Robot is super smart	.102	Robot could do my homework	.527
Robot makes plans for the weekend	.025	Robot goes grocery shopping	.046
Robot can keep a secret	.317	Robot would lie to me	.180
I would like to play a second quiz with the robot	.157	Robot is a slow thinker (scale)	.638
Robot would make a good study partner	1.000	Robot answered my questions quickly (scale)	.076
Robot goes on vacation	.129	Robot is successful at the quiz (scale)	.067
Robot answered my questions quickly	.008	Robot was a bit boring (scale)	.170

- Results for Wilcoxon test

Results for "Makes plans for the weekend" and "Does chores" must be disregarded for this test because of a violation of the symmetry assumption of the Wilcoxon test.

4.6.3 Final Questionnaire

A binomial, single-sample test was used to check for sufficiently biased response choices to reject the null (randomly distributed i.e. fifty-fifty) hypothesis.

Better at the quiz?	CF	13	.5	1.000
	NonCF	13	.5	
Do you find more human-like?	CF	23	.88	.000
	NonCF	3	.12	
Answers your questions more quickly?	CF	13	.5	1.000

	NonCF	13	.5	
Do you prefer (leuker)?	CF	18	.69	.076
	NonCF	8	.31	
Is smarter?	CF	11	.42	.557
	NonCF	15	.58	
Listened to you better while you were talking?	CF	14	.54	.845
	NonCF	12	.46	

4.6.4 Free Play Selection

A Wilcoxon related 2-sample test was conducted to analyze the relationship between free-play behaviour after interacting with the CF robot versus the Non-CF robot.

Ranks

		N	Mean Rank	Sum of Ranks
With Non-CF Robot - With CF Robot	Negative Ranks	3[a]	4.00	12.00
	Positive Ranks	4[b]	4.00	16.00
	Ties	19[c]		
	Total	26		

a. With Non-CF Robot < With CF Robot

b. With Non-CF Robot > With CF Robot

c. With Non-CF Robot = With CF Robot

Figure 10: A summary of the data. Children were slightly more likely to choose the quiz during free play after interacting with Non-CF robot

Test Statistics[a]

	With Non-CF Robot - With CF Robot
Z	-.378[b]
Asymp. Sig. (2-tailed)	.705

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Figure 11: Results of Wilcoxon test for significance.

It is very likely that the choice to continue with free play was not influenced by whether or not the children were interacting with a CF robot (p = 0.378).

4.6.5 Aliveness Ordering Task

The frequency of ordering the CF Robot as more alive than the Non-CF Robot were counted and tested for significance using a single-sample binomial test.

Robot	Instances Ranked Higher	Percentage Ranked Higher
CF	20	77 %
NonCF	6	23 %

The children most certainly tended to rank the CF robot as more alive than the NonCF robot (p=0.009).

The frequency with which the children ordered the CF robot as more alive than other objects were also compared. Three ranking tendencies were especially interesting:

Item	No. CF Ranked More Alive	
Plant	19	(73%)
Fish	12	(46%)
Dog	8	(31%)

Apart from two instances the two robots were placed immediately adjacent in the ranking. In the two exceptions the CF robot was ranked two places higher in the ranking than the NonCF robot, with the fish separating them in one instance and a dog in the other.

4.6.6 Photo Partner Preference

The frequency with which participants chose to have their photo taken with the CF robot was calculated and tested for the likelihood this frequency was non-random using a single-sample binomial test.

Robot	Times Chosen	Chosen as Percentage
CF	19	73 %
NonCF	7	27 %

It is very clear from this result that the children strongly preferred to have the CF robot as a photo companion, an implicit indication of preference, it is hypothesized (p = 0.028, binomial test).

4.6.7 Questions Asked in Dialog

During the dialog phase of the experiment the children were prompted to ask the robot a maximum of three questions. The children were not pressured into posing a second or third question. The number of instances where the child chose to ask more than one question was measured under the assumption that choosing to ask the robot more than one question may indicate a strong level of engagement.

Condition	Question 1	Question 2	Question 3
CF	26	22	12
NonCF	25	20	11

It is clear that the children did not feel any more or less inclined to ask the robot a second or third question based on whether or not the robot employed CFs.

4.6.8 Child Heart Rate and Heart Rate Variability

HR and HRV data were averaged for each section of the experiment and compared to baseline data obtained by averaging HR and HRV data from a three-minute relaxation period (see 'Measures' above for more detail). Technical difficulties led to only eighteen of twenty-six participant data being suitable for analysis. Detailed analysis was conducted for each child across conditions.

Condition	Section	Heart Rate	HR Variability
CF	Dialog	2.84	-20.39
CF	Quiz	2.24	-18.37
NonCF	Dialog	1.97	-28.48
NonCF	Quiz	1.25	-13.2

The results do not yield clear insights, however some interesting trends can be noted. Heart rates were more elevated for the CF robot than with the NonCF robot, likely indicating increased arousal and engagement. The largest drop in HRV corresponded to the the dialog portion of the NonCF condition with a resounding -28.48 decrease. The CF robot, however, caused, on average, a drop of only -20.39, nearly thirty percent smaller, when compared to a baseline. The results are made somewhat unclear when comparing the HRV for the quiz portions of the experiment with the CF arrangement actually registering a 18 % lower HPV than the NonCF version, contrary to expectations. The results to indicate a *trend* towards the CF-robot being more engaging and less stressful than the NonCF counterpart, but the results are made less clear by the exception mentioned above.

4.6.9 WoZ Stress Measurements

Length of Typed Responses The length of WoZ generated responses to children's questions during the dialog phase of the experiment were checked for length and compared. In both the CF and NonCF conditions the Wizard had an enforced minimum of time to type their response (seven seconds). Differences in the lengths of WoZ responses may indicate 1) a difference in perceived available time remaining and may indicate that the WoZ had a more accurate perception of time elapsed in one condition, and 2) That the WoZ experienced less stress, and thus lengthier respones, in one condition.

The average lengths of each WoZ response in both conditions are given below. Length is in number of characters.

Condition	**Question 1**	**Question 2**	**Question 3**
CF	46	40	41
NonCF	48	47	41

The results do not indicate any difference in time pressure.

WoZ Heart Rate Data The WoZ's heart rate and heart rate variability were recorded and analyzed to test the hypothesis that the WoZ would show elevated stress in the CF versus Non CF condition as a result of unfilled pauses. A decrease in heart rate variability has been shown to accompany conditions with acute time pressure (Nickel and Nachreiner, 2003) and cognitive load (Rowe et al., 1998).

The heart rate monitor sampled the WoZ's heart rate and heart rate variability every second during the course of each trial. The readings were taken for the dialog section of the experiment, averaged, then compared to a baseline created by averaging the HR and HRV readings for the WoZ for the last three minutes of the quiz (when they would have been sitting still for a prolonged period of time and well into the quiz). Technical difficulties led to only sixteen of twenty-six WoZ data being suitable for analysis.

Condition	HR	HRV
CF	3.24	-27.77
NonCF	4.7	-35.77

The results clearly indicate that WoZ heart rate was more heightened and heart-rate variability more lowered when controlling a robot not using CFs than when the robot did use CFs (when compared to a baseline). Heart rates increased on average 45% more for trials without CFs filling the pauses in conversation than they did in trials without CFs. Heart rate variability dropped an additional 28% when the gaps in conversation were without CFs. This strongly suggests that the WoZ felt less stress when controlling a CF-employing robot during a dialog with a child than when controlling a Non-CF employing robot.

4.7 Discussion

The results of this experiment clearly support the claim that conversational fillers are effective at mitigating the effects of robot response delay in interactions between children and speech-enabled anthropomorphic robots. The results show that not only do children prefer interacting with a robot with CF-filled gaps in conversation but that they do not seem, in turn, to view the robot as less intelligent or trustworthy as a result. The strongest measured effect was that robots employing CFs (with gestures) are overwhelmingly viewed as more human-like and alive than their non-CF employing counterparts. The children also reported that they felt the CF robot responded to their questions more quickly (despite the response times being identical across conditions).

Preference for the CF robot was exhibited in a two important ways. Firstly, when forced to choose one of the two robots as "leukst" (Dutch for 'best' or 'nicest'), 18 out of 26 children chose the CF robot. 73 % of children chose to have their photo taken with the CF robot, an implicit measure of likability.

The perceived increase in speed was observed in two out of three measures, with only the forced choice between robots as to the fastest responding showing no effect.

Almost all children (88%) indicated the CF robot to be the more human-like of the two robots while a significant effect was found between CF use and agreeing that the robot "made plans for the weekend", an indication of agency and humanness. Lastly, twenty out of twenty-six children ranked the CF robot as more alive than the non-CF robot in the aliveness ordering task, strongly indicating that the CF robot felt more alive than its counterpart (sig. = 0.009).

All these advantages came without any indication of the CF robot seemingly less intelligent than the Non-CF robot. The children's choice of which robot was better at the quiz were equally divided for each robot (13 for each). The CF robot and Non-CF robot were almost evenly chosen as the smarter of the two (11 versus 15) and neither was chosen more significantly more often as being a better listener (14 for the CF versus 12 for Non-CF robot). In fact, children were more likely to rate the CF-employing robot as more successful at the quiz despite that, on average, the two robots answered the same number of questions incorrectly.

Children's heart rates were slightly more elevated on average when interacting with the CF robot than with the Non-CF robot, although the limited number of trials made the difference statistically insignificant. They also did tend to ask the CF robot more questions during the dialog section of the experiment although these numbers are slight.

The iconic gestures appear to have successfully enhanced the effect of the CFs. As part of a concluding questionnaire filled in by the child participants in the current study, the children were asked to respond to the question of why they found one of the two robots more human-like. One child tellingly responded with "He moved more and scratched his head when he was thinking, this gave me the idea there was a person in front of me, which I liked". Or, another "Because he'll scratch his head when he has to think". It is interesting to note that not a single child chose the audible component of the CF as being their reason for selecting the CF robot as more human-like than the Non-CF robot (23 out of 26 children did).

Evidence for the effectiveness of using CFs and accompanying gestures to combat saturation comes from the ordering effects observed in the present work. Children that interacted with the CF robot in their first session took less notice of the loss of CFs in their second session. When children were asked to provide written answers to why they preferred one robot over the other, children that had the CFs in their first session referenced movement as part of their reasoning 6 out of 156 times. This number jumped almost three fold to 16 for responses from children who encountered the CFs in their second session. Additionally, only 77 % of the children from the former group rated the CF robot as more human-like compared to 100 % for the children that had the CFs in their second session. This elevated number for the latter group may be associated with a relief from saturation. Their first session with the Non-CF robotwas movement and behaviour poor and likely lead to saturation (i.e. boredom from lack of new behaviour). The second, with CFs, may have relieved this feeling of saturation by introducing more variety, thus amplifying the impression that the CF robot was more alive.

For the WoZ's part, heart-rate data seems to suggest that employing CFs can lower stress and perceived time pressure although, again, the relatively few number of sessions make hard conclusions difficult to draw. This effect was likely minimized by the WoZ's knowledge that a delay of a certain minimum duration was being enforced as part of the experiment's methodology. We would expect that the WoZ would feel greater time pressure and stress if they believed that it was their actions that caused the delay.

Overall the study strongly supports the case for using CFs to ameliorate issues from robot response delay in dialog. Children clearly felt that they contributed to the feeling that they were interacting with a human-like, living being that was more responsive, capable, and likable. The results should also assuage worries that using pensive fillers make the robot seem less intelligent, a worry not yet addressed by HRI research. Using fillers, especially those with accompanied by gestures, also create opportunities to combat saturation, that is, the loss of excitement observed in prolonged interactions.

5 Conclusion

It is not yet clear whether we will see truly autonomous, lifelike robots in the foreseeable future. Despite this uncertainty, however, it is evident that there is keen interest in creating robots that *seem* to possess genuine intelligence and aliveness, for research purposes or otherwise. There are several justifications for the charade. At this stage of AI research, robots can be seen as blunt tools that simply serve their intended function better when their interactors adopt richer or more human-like intentional stances to explain their behavior than if the robot exhibited a more mechanistic likeness. The situation can be seen as analogous to work being done to improve speech synthesizers to seem more human-like. The goal is not simply better quantifiable comprehension (number of words understood, for instance) but instead can be seen as the pursuit of more pleasant, meaningful, and effective artificial voice through human-voice and conversation mimicry.

Moral concerns must be addressed, however, when a human is covertly involved in controlling the robot's behaviour. In this case there is deception not only in the benign sense of the robot seeming to have emergent, human-like behaviour such as laughter or CFs, but also a deeper sense in which the interactor is being duped in a duplicitous manner. A child overwhelmed with the aliveness of a marionette in a puppet show will eventually, through maturity or through an overt breaking of the illusion, come to realize that puppets do not themselves hold the ability to walk and talk; an adult exposed to a robot in a WoZ setting, however, may never have the 'curtain lifted'. Apart from the immorality of deceitful practice, especially in scientific experimentation, this sets a level of expectation that robots will fail to meet, at least in the short term. Imagine the disappointment the purchaser of a Roomba, an 'autonomous' robotic vacuum, would have if they had interacted with a similar robot the day prior to their robot's acquisition that, through covert control from a WoZ, could have accepted voice commands or flawlessly navigated a complex landscape.

Nonetheless, there is a place for WoZ control of robots in the commercial or otherwise public realm. Should a purchaser of a robotic vacuum be made aware that a human may be called upon to remotely assist the robot if it gets stuck, for instance, they may well welcome the robot/human hybrid. It may well be that the holes left in the array of behaviours and abilities necessary for robots to become ubiquitous can be plugged, so to speak, by the (temporary) involvement of a remote operator. For instance, in one study (Glas et al., 2009) investigates the possibility of having a WoZ on hand to deal with unanticipated or overly complex situations arising with the simultaneous operation of four otherwise autonomous robots. In this situation the WoZ is automatically summoned when the robot fails to determine an appropriate course of action. There is usually a delay associated with this arrangement: a short pause when the operator is available but needs to be connected and a longer delay when the operator is busy with another robot. In each case findings from the present or related research is highly valuable in sustaining the IoA during such delays. CFs could be a very helpful tool in mitigating the delays associated with the types of teleoperation of robots that may well be an essential part of commercial (or otherwise widely available) robots of the future.

CFs have been shown to be of particular value in such scenarios. The intentional stance adopted by a robot's interactor is fickle, especially in adults. Unfilled response delay has presently been shown to not only be ill-favoured but indeed detrimental to the IoA when compared to robots with filled pauses. Robots employing CFs were seen as consistently more alive, autonomous, responsive, intelligent, and likeable. No small feat. Insofar as robot response delay is inevitable, most notably in WoZ situations, present or future, it is highly desirable to have well understood tools at hand to combat the negative effects that come along with undesirably lengthy response times. Until now it was simply assumed that the preference for CF-employing robots brought with it no unwanted side-effects. The results presented here suggest that not only are there no apparent negative side effects but that there are unforeseen positive ones. There is of course more work to be done before strong conclusions can be drawn. Nonetheless, the current results are indeed promising: one can not only expect that CFs make delayed response seem shorter, they can also anticipate that they increase the robot's perceived intelligence, aliveness, likeability, and autonomy without causing the (child) interactor to think less of its ability.

Employing CFs in conversational, anthropomorphic robots seems, from the present results, to also offer the opportunity to combat saturation (i.e. boredom from prolonged exposure to unchanging behaviour) without deviating from the planned interaction (so long as the interaction involves a two-way conversation). An HRI designer can introduce varied CFs that introduce novel behaviours throughout a prolonged interaction without fear that the CFs are detrimental to intelligence the (child) interactor ascribes as part of its intentional stance. Put another way, CFs offer an opportunity for a robot to behave in a more and more natural, human-like, and entertaining way without, it is suggested presently, negatively affecting other parts of a child's IoA (indeed *enhancing* them).

The plethora of humanoid robots available today make clear the desire to create robots that illicit an intentional stance similar to those we adopt for fellow humans. While AI struggles to create robots that can live up to this demanding IoA, it is important that we both hunt for ways to endow our robots with (non-emergent but pre-programmed) behaviours that make them seem more human-like and autonomous but also fully understand the subtleties that such 'tricks' bring along with them. Until now CFs have been widely used to ameliorate response delays in a human-like way; what was lacking was a comprehensive account of the full ramifications their use has on a (child) interactor's illusion of agency. The present work strongly suggests that CFs do indeed help mitigate the feeling that a robot is unresponsive during delays in dialog, but also bring with their use a slew of positive effects.

6 Future Work

This work does reveal interesting avenues for future study. For instance, CFs in natural language are often used to demarcate false starts, that is, marking an utterer's intention to restart an incomplete utterance (often with corrected information or intention) (Tree, 1995), for instance, the "um" in "I wanted to, um, I went to the mall". In this instance a CF may be used not (only) to fill an unwanted gap in a conversation but also to signal that a correction to vocalized misinformation is going to replace the 'false start' contained in the interrupted sentence. This may prove to be another tool particularly valuable to WoZ settings as it is commonplace to trigger an incorrect 'canned' or preprepared robot response during dialog, for instance.

Research into delays in webpage load times, on the one hand, and naturalistic uses of CFs to signal upcoming important information in dialog, also suggests that CFs may be useful in providing what are called "deterministically predictive" response delays. (Thomaschke and Haering, 2014) suggests that if response delays are inevitable, it is better to have the delays of a standard, repeated length such that the length of the delay predicts the type of information that will follow. It has been shown to speed the overall interaction to such an extent that prolonging a delay to make it predictive (i.e. indicative of the type of information that will follow) is actually beneficial to the overall interaction duration. Furthermore, the authors point out that in cases where the web page is meant to provide hard-to-obtain information, such as scouring the web for inexpensive and convenient airline tickets, websites purposely introduce lengthy delays so that the users feel that an exhaustive, complex search has taken place. Relatedly, CFs have been shown to be more present before the vocalization of important information, indicating that CFs act as a sort of signal. Listeners were more able to identify words from recorded speech when the words were proceeded by the fillers "um" and "ah", for instance (Pfeifer and Bickmore, 2009).

Collectively these studies suggest another potential use for CFs. Their duration could be tactically chosen to signal the type of response the robot will make, for instance. Relatedly, they could be used not to fill a response delay but instead as a marker for upcoming important information. These potential uses have not yet been explored but are avenues that could, it seems, yield interesting and useful results for HRI.

A final, important consideration is alluded to in the background section of this document. Children and adults have very different attitudes, expectations, and beliefs towards robots. It is an important question whether the results presented here generalize to older interactors, that is, whether adults would see a robot employing CFs as speedier, more alive, and capable. It could be the case that, for instance, an adult interactor would rate a CF-employing robot as less intelligent since they may have stronger attitudes towards people (or humanoid robots) showing pensiveness before responding. While purely hypothetical, this and other unanswered concerns should be addressed before CFs are integrated into a robot intended to interact with non-adolescent interactors.

The research presented here makes important contributions to HRI research. It has been shown that CFs are useful at filling unwanted response delays in WoZ scenarios from both the child participants' points of view and from the WoZ's. The subtleties of the change in perception arising from using CFs in these situations have been elucidated and have been shown to be a effective at not only ameliorating the negative effects arising from unwanted robot response delay but also in making the robot more believable as a human-like agent. Some related future research avenues have been identified and there is reason to expect these pursuits to be contributory to the field of HRI and AI as a whole.

References

Aucouturier, J.-J. and Ikegami, T. (2009). The illusion of agency: Two engineering approaches to compromise autonomy and reactivity in an artificial system. *Adaptive Behavior*, 17(5):402–420.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125.

Belpaeme, T., Baxter, P. E., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., Racioppa, S., Kruijff-Korbayová, I., Athanasopoulos, G., Enescu, V., et al. (2012). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2):33–53.

Blanson Henkemans, O. A., Bierman, B. P., Janssen, J., Neerincx, M. A., Looije, R., van der Bosch, H., and van der Giessen, J. A. (2013). Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Patient education and counseling*, 92(2):174–181.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech*, 44(2):123–147.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., and Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 708–713. IEEE.

Breazeal, C. and Scassellati, B. (1999). How to build robots that make friends and influence people. In *Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on*, volume 2, pages 858–863. IEEE.

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Fernaeus, Y., Jacobsson, M., Ljungblad, S., and Holmquist, L. E. (2009). Are we living in a robot cargo cult? In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 279–280. ACM.

Fischer, K. (2006). The role of users preconceptions in talking to computers and robots. In *Proceedings of the Workshop on How People Talk to*

Computers, Robots, and other Artificial Communication Partners, pages 112–130.

Fox, R. and McDaniel, C. (1982). The perception of biological motion by human infants. *Science*, 218(4571):486–487.

Fraser, N. M. and Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language*, 5(1):81–99.

Friedman, B., Kahn Jr, P. H., and Hagman, J. (2003). Hardware companions?: What online aibo discussion forums reveal about the human-robotic relationship. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 273–280. ACM.

Glas, D. F., Kanda, T., Ishiguro, H., and Hagita, N. (2009). Field trial for simultaneous teleoperation of mobile social robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 149–156. ACM.

Glas, D. F., Kanda, T., Ishiguro, H., and Hagita, N. (2012). Temporal awareness in teleoperation of conversational robots. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 42(4):905–919.

Harris, J. and Sharlin, E. (2011). Exploring the affect of abstract motion in social human-robot interaction. In *RO-MAN, 2011 IEEE*, pages 441–448. IEEE.

Kahn Jr, P. H., Friedman, B., Perez-Granados, D. R., and Freier, N. G. (2004). Robotic pets in the lives of preschool children. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1449–1452. ACM.

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-computer interaction*, 19(1):61–84.

Kanda, T. and Ishiguro, H. (2005). Communication robots for elementary schools. In *Proc. AISB05 Symposium Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, pages 54–63.

Kanda, T., Ishiguro, H., Imai, M., and Ono, T. (2003). Body movement analysis of human-robot interaction. In *IJCAI*, pages 177–182. Citeseer.

Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H., and Anzai, Y. (2007). A humanoid robot that pretends to listen to route guidance from a human. *Autonomous Robots*, 22(1):87–100.

Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., and Hagita, N. (2009). An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 173–180. ACM.

Komatsu, T., Kurosawa, R., and Yamada, S. (2012). How does the difference between users expectations and perceptions about a robotic agent affect their behavior? *International Journal of Social Robotics*, 4(2):109–116.

Maulsby, D., Greenberg, S., and Mander, R. (1993). Prototyping an intelligent agent through wizard of oz. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 277–284. ACM.

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental psychology*, 31(5):838.

Midden, C. and Ham, J. (2012). The illusion of agency: the influence of the agency of an artificial agent on its persuasive power. In *Persuasive Technology. Design for Health and Safety*, pages 90–99. Springer.

Nalin, M., Bergamini, L., Giusti, A., Baroni, I., and Sanna, A. (2011). Childrens perception of a robotic companion in a mildly constrained setting. In *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop) Proceedings*.

Nickel, P. and Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-hz component of heart rate variability as an indicator of mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(4):575–590.

Okita, S. Y. and Schwartz, D. L. (2006). Young children's understanding of animacy and entertainment robots. *International Journal of Humanoid Robotics*, 3(03):393–412.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist*, 17(11):776.

Pfeifer, L. M. and Bickmore, T. (2009). Should agents speak like, um, humans? the use of conversational fillers by virtual agents. In *Intelligent Virtual Agents*, pages 460–466. Springer.

Read, R. and Belpaeme, T. (2014). Non-linguistic utterances should be used alongside language, rather than on their own or as a replacement. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 276–277. ACM.

Riek, L. D. (2012). Wizard of oz studies in hri: A systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1).

Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., and Olsson, L. (2005). Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 716–722. IEEE.

Ros, R., Nalin, M., Wood, R., Baxter, P., Looije, R., Demiris, Y., Belpaeme, T., Giusti, A., and Pozzi, C. (2011). Child-robot interaction in the wild: advice to the aspiring experimenter. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 335–342. ACM.

Rowe, D. W., Sibert, J., and Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 480–487. ACM Press/Addison-Wesley Publishing Co.

Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N. (2008). How quickly should communication robots respond? In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 153–160. IEEE.

Tanaka, F., Cicourel, A., and Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, 104(46):17954–17958.

Thomaschke, R. and Haering, C. (2014). Predictivity of system delays shortens human response time. *International Journal of Human-Computer Studies*, 72(3):358–365.

Tielman, M., Neerincx, M., Meyer, J.-J., and Looije, R. (2014a). Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014*

ACM/IEEE international conference on Human-robot interaction, pages 407–414. ACM.

Tielman, M., Neerincx, M., Meyer, J.-J., and Looije, R. (2014b). Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14)*, pages 407–414.

Tojo, T., Matsusaka, Y., Ishii, T., and Kobayashi, T. (2000). A conversational robot utilizing facial and body expressions. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 2, pages 858–863. IEEE.

Tree, J. E. F. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of memory and language*, 34(6):709–738.

Turkle, S., Breazeal, C., Dasté, O., and Scassellati, B. (2006). Encounters with kismet and cog: Children respond to relational artifacts. *Digital Media: Transformations in Human Communication*, pages 1–20.

Ward, N. (2000). The challenge of non-lexical speech sounds.

Ward, N. (2006). Non-lexical conversational sounds in american english. *Pragmatics & Cognition*, 14(1).

A Supporting Materials

A.1 Nao Robot

The Nao robot is a 58cm tall, linux based humanoid robot developed by the French company Aldebaran. It is equipped with a wide array of sensors and actuators: 4 microphones, 2 cameras, a gyroscope, an accelerometer, 2 speakers, and 2 infrared and sonar sensors. It is equipped with WiFi and Ethernet and runs a custom built form of Linux called "Naoqi", an operating system specifically designed for the robot. It is quite friendly and non-threatening in appearance ad is particularly well suited for child-robot interaction (Nalin et al., 2011). The robot is capable of walking, crouching, grasping, and, to a large extent, self-balancing. Behaviors for the Nao can be easily created using a GUI-based software package from Aldebaran called "Choregraphe".

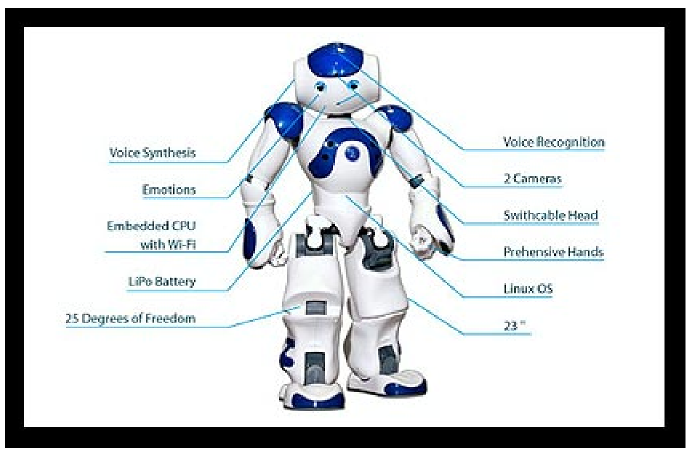


Figure 12: The various capabilities of the Nao, courtesy of Aldebaran

A.2 Seesaw

The 'Seesaw' is a simple lever/fulcrum device designed to hold a ten-inch tablet computer at a roughly forty-five degree angle in one of two opposing positions. The contraption stands roughly 30 cm tall and affixes the tablet PC with four l-shaped, rubber-protected sticks. Once attached, a simple lift at the tablet's edge allows the tablet to be switched to face one of two users facing each other with the tablet in the middle. This simple device, custom built at TNO Soesterberg, allows for a pleasant turn-taking experience at the center of the quiz activity between the participant and the robot. The tablet automatically flips its screen to match its orientation (so text is not inverted).

A.3 Video Camera

A single 1080p video camera was used to capture every session for video analysis. It was placed atop a standard tripod roughly two meters from the child and was made to have the child's face and tablet in its view. The participants signed a consent form permitting their image to be captured and were reminded at the outset of each session that they would be filmed. The video was fed to a laptop running Noldus's Observer XT software for real-time video annotation performed by one of the experimenters.

A.4 Neutral Movements

From the moment when the robot stood at the beginning of the session to the time it crouched at its conclusion, when the robot was not performing an overriding movement (CF, flipping the tablet) the robot would automatically perform what we dubbed "neutral movements". These movements were semi-random, slight movements of the robot's arms, head, and stance that were intended to make the robot seem more lifelike. The robot performed these movements when talking in both conditions (i.e. having the robot speak did not pause these movements).

B Robot Response Delay Times

During the quiz section of the experiment the delays purposefully introduced between the conclusion of the child asking the robot a question and the robot's response were fixed. These times were identical for each session (with CFs filling the delays in one of the two sessions). These times were, in milliseconds:

7600,5600,8400,6400,8000,7200

C Questionnaires

Post-Trial Questionnaire

1. **Ik wil graag vriendjes zijn met Charlie**
   - ○ Nee ○ Misschien ○ Ja
2. **Charlie lijkt me super slim**
   - ○ Nee ○ Misschien ○ Ja
3. **Charlie maakt plannen voor het weekend**
   - ○ Nee ○ Misschien ○ Ja
4. **Ik denk dat Charlie een geheim kan bewaren**
   - ○ Nee ○ Misschien ○ Ja
5. **Ik wil graag nog een tweede quiz met Charlie spelen**
   - ○ Nee ○ Misschien ○ Ja
6. **Charlie zou een goede studievriend zijn**
   - ○ Nee ○ Misschien ○ Ja
7. **Charlie gaat op vakantie**
   - ○ Nee ○ Misschien ○ Ja
8. **Charlie heeft snel op mijn vragen geantwoord**
   - ○ Nee ○ Misschien ○ Ja
9. **Ik vertrouw Charlie**
   - ○ Nee ○ Misschien ○ Ja
10. **Ik denk dat Charlie mijn huiswerk zou kunnen maken**
    - ○ Nee ○ Misschien ○ Ja
11. **Charlie gaat boodschappen doen**
    - ○ Nee ○ Misschien ○ Ja
12. **Ik denk dat Charlie zou liegen tegen mij**
    - ○ Nee ○ Misschien ○ Ja

13. Charlie is een trage denker

Minder Meer

14. Charlie heeft snel op mijn vragen geantwoord

Minder Meer

15. Charlie doet de quiz met succes

Minder Meer

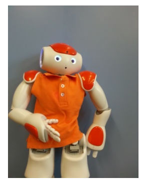
16. Charlie was een beetje saai

Minder Meer

Final Questionnaire

1. **Welke robot is beter in de quiz?**

- Charlie

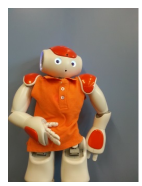


- Robin

Waarom?

2. **Welke robot vind je het meest menselijk ?**

- Charlie



- Robin

Waarom?

3. **Welke robot antwoordt het snelst?**

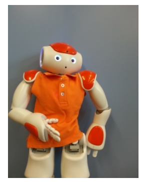
- Charlie

- Robin

Waarom?

4. **Welke robot vind je het leukst?**

○ Charlie

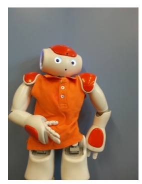


○ Robin

Waarom?

5. **Welke robot is slimmer ?**

○ Charlie



○ Robin

Waarom?

6. **Welke robot luisterde beter terwijl jij aan het praten was?**

○ Charlie

○ Robin

Waarom?

D Responses to 'Why?' Questions

(Dutch Original / English Translation)

	Welke Robot is beter in de quiz?	Which robot is better at the quiz?
1	[No Answer Provided]	[No Answer Provided]
2	Weet ik niet	I don't know
3	Hij is sneller met nadenken en heeft meer antwoorden goed	Robot thinks faster and has more answers correct
4	Omdat ik hem beter kon verstaan	He was easier to understand
5	Die geeft sneller antwoord	he responds more quickly
6	Hij antwoord wat sneller	he responds more quickly
7	Robin had alle vragen in 1 keer goed	Robin answered correctly at the first try for every question.
8	Die had bij mij meer vragen goed	He answered correctly more often
9	Hij is sneller	He is quicker
10	Antwoordde sneller	He answered more quickly
11	Omdat ik de vraag of het antwoord beter kon horen	I could hear him more easily
12	Hij koos steeds het goede antwoord	He selected the right answer each time
13	[No Answer Provided]	[No Answer Provided]
14	Het lijkt echt of hij nadenkt	It really looks like he thinks
15	[No Answer Provided]	[No Answer Provided]
16	Hij antwoordde eerder op de vragen	He responds more quickly
17	[No Answer Provided]	[No Answer Provided]
18	Hij antwoord wat sneller maar ik weet niet meer of hij wel beter dan charlie had gedaan	[No Answer Provided]
19	Want die reageert wat sneller	Because he responds more quickly
20	Hij was wat rustiger	He was a bit more quiet
21	Hij beweegt meer dat is fijner	He moves more, which is nicer.
22	Omdat hij sneller antwoord gaf	Because he answered more quickly
23	Hij heeft wat sneller vragen goed	He answers correctly more often
24	Hij had meer vragen goed	He answered correctly more often
25	Hij had meer vragen goed	He answered correctly more often
26	Dat kan ik niet zo goed uitleggen	I can't explain

	Welke robot vind je het meest menselijk?	Which robot do you find most human-like?
1	[No Answer Provided]	[No Answer Provided]
2	[No Answer Provided]	[No Answer Provided]
3	Hij beweegt gewoon meer menselijker	He moves in a more human way
4	Want hij is heel menselijk	Because he is very human
5	Beweegt wat meer	He moves more
6	Omdat hij strepen op zijn t-shirt heeft	Because he has a striped t-shirt
7	Hij beweegt het meest	He moves the most
8	Die praat wat langzamer	He talks a bit more slowly
9	[No Answer Provided]	[No Answer Provided]
10	Beweegt meer	He moves more
11	Hij beweegt meer	He moves more
12	Omdat hij ook soms fouten maakt	Because he sometimes makes mistakes
13	Hij beweegt veel meer	He moves a lot more
14	Hij kan ook aan zijn hoofd voelen en meer dingen	He can scratch his head and do other things
15	Hij had wel meer mens in zich	He had a bit more human in him
16	Hij kan staan	He can stand up
17	Hij beweegt best veel	He moves quite a lot
18	Die beweegt wat meer en als hij nadenkt krabt hij som ook op zijn hoofd	He moves more and scratches his head when he thinks
19	Want die gaat als hij moet denken bijvoorbeeld met zjin handen op zijn hoofd krabbelen of zoiets	Because he'll scratch his head when he has to think
20	Bij de quiz leek hij ook echt een mens	He looked human during the quiz
21	Omdat hij veel beweegt, Charlie blijft een beetje stilstaan	Because he moves. Robin stands still a lot.
22	Omdat je ziet dat hij nadenkt over de vragen	Because you can tell he's thinking about the questions
23	Hij sosm zijn hoofd krabt	Because he sometimes scratches his head

24	Hij bewoog meer en krabte ook op zijn hoofd als hij nadacht (dit gaf mij wel het gevoel dat er iemand voor mij stond (dit was wel fijn))	He moved more and scratched his head when he was thinking, this gave me the idea there was a person in front of me, which I liked.
25	Hij doet iets met zijn armen	He's doing something with his arms
26	Hij beweegt meer	He moves more
	Welke robot antwoordt het snelst?	**Which robot answers the quickest?**
1	[No Answer Provided]	[No Answer Provided]
2	[No Answer Provided]	[No Answer Provided]
3	Hij is slimmer	He is smarter
4	Dat weet ik niet	I don't know
5	Als ik een vraag stel geeft hij gelijk antwoord	He responds immediately after my question
6	Hij hoort mij denk ik beter	I think he hears me better
7	[No Answer Provided]	[No Answer Provided]
8	Geen idee	No idea
9	[No Answer Provided]	[No Answer Provided]
10	Het waren misschien makkelijkere vragen	The questions may have been easier
11	Als ik de vraag stelde dacht hij 5 seconden na en gaf het antwoord	When I asked the question he thought for 5 seconds before answering
12	Omdat hij het sneller wist	Because he knew the answer more quickly
13	Hij is een beetje sneller	He's a little faster
14	"He just does"	He just does
15	Geen een, even snel	Neither
16	Gewoon	Just because
17	[No Answer Provided]	[No Answer Provided]
18	Omdat hij sneller antwoord	because he answers more quickly
19	Robin, die was wat langer aan het nadenken	Robin, because he was thinking for a longer period of time
20	Hij luistert goed	He's a good listener
21	Hij antwoordt gewoon sneller	He just answers more quickly
22	Omdat die meer beweegt	Because he moves more
23	Weet het antwoord sneller	Because he knows the answer more quickly

24	Robin krabte eerst nog op zijn hoofd	Robin scratched his head before answering
25	Weet ik niet	I don't know
26	Dat kan ik ook niet zo goed uitleggen	I can't explain
	Welke robot vind je het leukst?	**Which robot do you to be the best?**
1	[No Answer Provided]	[No Answer Provided]
2	[No Answer Provided]	[No Answer Provided]
3	Hij reageert blijer	He responds in a more cheerful way
4	Want hij is schattiger	Because he is cuter
5	Vraagt wat meer	He asks more questions
6	Omdat hij toch anders is dan Robin	Because he is different from Robin
7	[No Answer Provided]	[No Answer Provided]
8	Ik vind ze eigenlijk allebei even leuk	I like them both equally
9	Hij is menselijker	He is more human
10	Hij was erg lief	He was very sweet
11	Hij was soms grappiger	He was funnier sometimes
12	Hij is leuk	He's more fun
13	Hij doet bijna alles een beetje beter	He does almost everything a little better
14	Ik vind hem meer menselijk	I think he's more human
15	Allebij, hij was heel erg leuk aan het nadenken	Both, the thinking was cute
16	Omdat hij veel beweegt	Because he moves a lot
17	[No Answer Provided]	[No Answer Provided]
18	Ik vind hem wat echter lijken	I think he looks more real
19	Hij heel erg menselijk probeerde te doen, en hij was wel lief	Because he tried to be very human like and he was kind of sweet
20	Hij was super aardig	He was very nice
21	Hij is gezelliger vind ik maar Robin is ook leuk	Charlie is nicer, but Robin is also nice
22	Omdat hij beweegt en dat is grappigq	Because he moves, which is funny
23	Omdat hij meer menselijk is	because he's more human
24	Die leek veel menselijker	He looked more human
25	Hij beweegt wat meer	He moved more
26	[No Answer Provided]	[No Answer Provided]

	Welke robot is slimmer?	**Which robot is smarter?**
1	[No Answer Provided]	[No Answer Provided]
2	[No Answer Provided]	[No Answer Provided]
3	Hij heeft meer antwoorden goed en iseen snelle denker	He answers correctly more often and is a quicker thinker
4	Omdat hij de meeste antwoorden goed had	Because he answered correctly most often
5	Wist meer vragen	He knew more answers
6	Hij had meer vragen goed op de quiz	He had more right answers
7	Alle antwoorden in 1 keer goed	Robin answered each question correctly the first try
8	Die had meer vragen goed	He responded correctly more often
9	Hij beantwoordde meer vragen goed	He answered correctly more often
10	Had eerder de vragen goed	He anwered correctly more quickly
11	Hij had de meeste antwoorden goed	Hij answered correctly most often
12	Hij koos steeds het goede antwoord	He chose the right answer all the time
13	[No Answer Provided]	[No Answer Provided]
14	Hij antwoordde sneller	He answered more quickly
15	Hij zei van de 3 er 2 goed	He answered correctly 2/3 times
16	Hij had meer vragen goed	He answered correctly more often
17	hij antwoordde meer goed	He answered correctly more often
18	Ik weet het niet zeker ik denk het gewoon	I'm not sure, that's just what I think
19	Omdat hij bij de quiz best wel snel antwoordde	Because he answered kind of quickly
20	Hij antwoordde sneller	He answered more quickly
21	Robin moest 4 keer raden voor hij het antwoord wist	Robin had to guess 4 times before answering correctly
22	Omdat hij sneller antwoord gaf	because he answered more quickly
23	Omdat hij sneller het antwoord gaf	Because he answers more quickly
24	Die had meer antwoorden goed	He answered correctly more often
25	Hij had het meeste antwoorden goed	He answered correctly more often
26	Het lijkt dat hij slimmer is	He seems more clever
	Welke robot luisterde beter terwijl jij aan het praten was?	**Which robot listened better while you were talking?**
1	[No Answer Provided]	[No Answer Provided]
2	[No Answer Provided]	[No Answer Provided]
3	Alle bij	Both

4	Omdat hij heel snel antwoord gaf	Because he answered quickly
5	Keek meer naar mij	He looked at me more often
6	Bij Charlie sprak hij mijn naam uit als sushi	Charlie pronounced my name as sushi
7	Hij wachtte ook tot je uigepraat bent en dan zegt hij iets	He waited until you finished talking before responding
8	[No Answer Provided]	[No Answer Provided]
9	[No Answer Provided]	[No Answer Provided]
10	Ik stelde hem meer vragen	I asked him more questions
11	Hij was stiller	He was more quiet
12	Weet ik niet	I don't know
13	Hij geeft ook antwoord	He answered too
14	"He just does"	He just does
15	Hij was meteen stil	He was silent immediately
16	Ze luisterden even goed	They both listened equally well
17	[No Answer Provided]	[No Answer Provided]
18	Zij bewoog dan wat minder	She moved a little less
19	Omdat hij je aankeek wat meer	Because he looked at you a bit more
20	Hij zat je aan te kijken terwijl je praat	He was looking at you while you were talking
21	Want Robin zat een beetje met z'n vingers te friemelen, dat was vervelend.	Because Robin moved his fingerrs, which was annoying
22	Omdat ik een keer B zei en Charlie dacht dat ik D zei	Because he thought I said B, while Charlie thought I said D
23	Ze wachtte tot je uitgepraat bent	She waited until you were done talking
24	Het leek echt of hij nadacht over wat ik zei	It really looked like he thought about what I was saying
25	Hij zei wat meer	He talked more
26	Hij hoort beter en sneller denk ik	He hears better and more quickly, I think