# A Farewell to Arms: Non-verbal Communication for Non-humanoid Robots*

**Aaron G. Cass**
Union College
Schenectady, NY, USA
cassa@union.edu

**Kristina Striegnitz**
Union College
Schenectady, NY, USA
striegnk@union.edu

**Nick Webb**
Union College
Schenectady, NY, USA
webbn@union.edu

## Abstract

Human-robot interactions situated in a dynamic environment create a unique mix of challenges for conversational systems. We argue that, on the one hand, NLG can contribute to addressing these challenges and that, on the other hand, they pose interesting research problems for NLG. To illustrate our position we describe our research on non-humanoid robots using non-verbal signals to support communication.

## 1 Introduction

Our research is about interaction strategies for robots who have to approach and communicate with strangers in busy public spaces (Cass et al., 2015, 2018). For example, in one of our target scenarios a delivery robot in a busy academic building on a college campus has to solicit help to operate the elevator from humans passing by. In another scenario a robot is recruiting survey participants in a shopping mall. In order to develop solutions that will work in a real-world deployment, we collect data and study human-robot interaction not just in laboratory experiments but also in field studies conducted in the wild.

In these field studies we have encountered challenges that are traditionally not addressed by the natural language generation (NLG) pipeline. However, we would like to argue that an NLG system aware of these issues can contribute to a better solution and that they also pose interesting research problems for NLG.

In particular, the following two sources of challenges have stood out to us. First, the robot is situated in a dynamic environment with human interaction partners that can act while the robot is speaking or planning an utterance. As in other situated communication tasks (Koller et al., 2010; Smith et al., 2011) the timing of the robot's utterances is important. For fluent interactions the robot needs to monitor the human's actions and changes in the environment and react to them in a timely manner, potentially by interrupting itself or modifying an utterance mid-stream (Clark and Krych, 2004).

Second, many environmental factors may hinder communication and are not controllable by us or the robot. For example, in a busy public space the background noise level may be high, making it hard for people to hear the robot. People may be passing by and even in between the robot and the addressee. The robot will encounter many different reactions from addressees; some will be surprised, scared, or embarrassed to interact with it.

One approach to these challenges would be to solve these issues first in order to create a situation where a "traditional" NLG pipeline, based on NLG for text generation, can be used optimally. For example, we could try to develop a module that perfectly times utterances, make sure to adjust the audio level to always be above the environmental noise level, and only communicate with addressees that are directly in front of the robot. However, these goals may be impossible to achieve. For example, while it makes sense to optimize the timing of utterances, most contributing factors are out of our control, so that the robot will always have to be prepared to deal with unexpected actions by the human addressee, changes in the environment, or network delays. Furthermore, this approach may lead to suboptimal results. For instance, if the robot only communicates with people if they are positioned right in front of it, in a busy space with people passing through, many opportunities for interaction may be lost.

Therefore, we believe that NLG should be

*Position paper presented at the workshop on natural language generation for human-robot communication at INLG 2018.



*Proceedings of the Workshop on NLG for Human–Robot Interaction*, pages 22–26,
Tilburg, The Netherlands, November 8, 2018. 

aware of these issues and can contribute to a solution. For example: An incremental NLG module may be able to better time utterances and react to unexpected changes (Allen et al., 2001; Buschmeier et al., 2012). When the environment is noisy or the robot is far away from the addressee, generating shorter utterances using simpler words and complementing speech with non-verbal signals might be more effective. Previous work has explored the problem of adapting the form and content of generated utterances to situational constraints (e.g. Jokinen and Wilcock, 2003; Walker et al., 2007; Rieser et al., 2014), but typically not in the context of human-robot interaction.

In order to illustrate our position, we will describe some results and observations from our ongoing research on making human-robot communication more robust using non-verbal signals. A lot of work has been done on generating non-verbal signals, like gestures, facial expressions, and posture for animated characters (known as embodied conversational agents or virtual humans). Some of this work has been transferred to humanoid robots. However, because of our application scenario, the use of humanoid robots is not practical for us. We need robots that are tall enough to interact with standing humans and that are not too expensive to be deployed in a busy public space. We work with robots that have a wheeled base and a mounted screen (see Figure 1).

The research challenge is, therefore, to find out what non-verbal signals are effective communicative devices for these non-humanoid robots. These signals may mimic human behaviors, or they may be visual metaphors that express the robot's intentions in a way that is not modeling realistic human behavior, similar to the way comics express a character's movement or emotions.

## 2 Related Work

People accompany their speech with non-verbal signals, which support and add to the content of the speech and which help manage the dialog. For example, iconic hand gestures may depict some features of an object or event being described (McNeill, 1992), eye gaze plays an important role in regulating turn-taking in dialog (Kendon, 1967), and facial displays express a speaker's emotions (Ekman and Friesen, 2003) but also serve pragmatic functions that help organize the dialog (Chovil, 1991).

Embodied conversational agents (ECAs) or virtual humans are animated characters that engage with humans in a dialog using both verbal and non-verbal communication (Cassell et al., 2000). Typical research in this area closely analyzes human non-verbal behavior and aims to model these behaviors in the animated character.

Some of this work on generating non-verbal behaviors for animated conversational characters has been transferred to physical humanoid robots. Salem et al. (2012) and Hasegawa et al. (2010) use gesture generation strategies developed for ECAs on humanoid robots. Breazeal (2000) presents a robot with a simple cartoonish face that can express emotions and interaction cues. Most expressions are modeled on human facial expressions. But the robot can also use its non-human, animal-like ears to indicate arousal and attention.

While Breazeal's (2000) work shows that even with humanoid robots going beyond the normal human repertoire of non-verbal signals can be beneficial, non-humanoid robots often are not capable of mimicking human non-verbal behaviors. It is therefore essential to identify what behaviors of non-humanoid robots can easily be interpreted by humans (Cha et al., 2018). Recent work has, for example, explored the interpretability of robot arm movements (Dragan et al., 2013). In a study that is most similar to our research, Cha and Matarić (2016) have shown that a service robot can use light and sound signals to indicate that it needs help and to communicate levels of urgency.

## 3 Experiments Exploring Non-verbal Signals for Non-humanoid robots

We describe three studies we have carried out or are currently conducting to explore how non-verbal behaviors can contribute to communication between humans and non-humanoid robots. In these studies we explore non-verbal robot behaviors modeled on human behaviors as well as robot behaviors designed to communicate metaphorically through movement.

The two robots we have used for this work, SARAH and VALERIE, both have a mobile base, a screen on which a simple cartoon face can be displayed, and a suite of cameras and depth sensors (VALERIE is shown in Figure 1). Importantly, the robots have a non-humanoid form, lacking the typical mechanisms for human non-verbal expression. Our experiments are conducted using

Figure 1: VALERIE

a Wizard of Oz (WoZ) protocol, in which a human wizard remotely controls the robot unbeknownst to the participants. The wizard interface provides a set of pre-planned behaviors the wizard can initiate, as well as lower-level controls for the robot.

### 3.1 Robot eye gaze to support reference

In this ongoing study we look at whether humans use our robot's eye gaze to resolve referring expressions. Hanna and Brennan (2007) found that humans use a human instruction giver's eye gaze to distinguish an object being described from its similar looking distractors. We replicated their experiment, in the laboratory, with VALERIE taking the instruction giver's place.

Participants stood opposite VALERIE with a table between them. On the table was a sequence of colored shapes, each of which also had a number of black dots. Some layouts contained distractor pairs, which are shapes of the same color and form, but with a different number of dots. VALERIE gave instructions of the form *"Press the button corresponding to the blue triangle with three dots"*, while either only moving her mouth or, additionally, accompanying the instruction by a movement of the pupils in the direction of the target shape. A preliminary analysis of the data suggests that VALERIE's eye gaze helps participants pick out the right target more quickly in situations where the layout contains a distractor shape that is sufficiently far away from the target shape that it can be distinguished by eye gaze.

This shows that the participants do indeed interpret the robot's eye gaze and use it to guide their own behavior. From an NLG point of view, the generation of eye gaze is interesting because eye gaze has to be coordinated with the natural language utterance it accompanies, while also producing natural looking eye movements.

*Limitations and future work:* This study was done in a laboratory environment using a repetitive and unrealistic task. We plan to conduct a follow up study that tests the effectiveness of robot eye gaze as a communicative device in the wild.

### 3.2 Robot body movement and orientation to attract attention and initiate interactions

In this experiment in the wild, the robot behavior was designed to (very crudely) mimic the behaviors humans might use to initiate an interaction with a passer-by in a busy public space. SARAH was stationed in a popular hallway. She would greet people ("Hello! Can you please help me?") either while standing still or accompanied by a rotational movement that followed the subject we wished to engage. People who approached SARAH were then asked to press a specific number on a keypad.

We collected video data of 14 one-hour sessions over the course of 5 weeks. In total, 1658 people passed by SARAH. Of those, only 714 engaged with her in any way, including just looking at her. Of the 714, 108 completed our task. We found that movement of the robot statistically significantly increased how many people looked at SARAH (64% of passers-by noticed the still robot, 88% the moving robot), but not the number of completed tasks (6.4% in the non-moving condition, 6.7% in the moving condition). Given a 30% increase in the number of people who notice SARAH, we expected a similar increase in the number of people who stop to interact with her.

A closer analysis of our video data indicates that technical issues with the WoZ interface (which we plan to address in follow-up experiments) as well as issues related to SARAH's communicative behavior may be the reason for why the increased attention did not lead to more successful interactions. First, it seems that SARAH's intentions weren't always clear and, second, several people in the study acted surprised or scared of SARAH or embarrassed to interact with her. Both issues point toward a need for better communicative non-verbal behaviors to convey the robot's intentions and to lessen people's apprehension.

As with eye gaze, these non-verbal behaviors have to be planned and coordinated with the robot's natural language utterances. An additional challenge is that the signals we are exploring are complex, involving eye gaze, facial expressions, and different kinds of movement. Furthermore,

the optimal choice of non-verbal signals and form or natural language utterance may depend on aspects of the environment, such as how busy and noisy it is or how far away the addressee is. The NLG system planning these utterances will have to be able to coordinate diverse types of communicative signals and to adapt to the current situation.

*Future work:* In our current work, we are studying verbal and non-verbal behaviors that allow the robot to better signal its wish to interact (e.g. moving toward the selected addressee, facial expressions to indicate a need for help and a wish to engage). This exploration is guided by what is known about human behaviors in similar situations (Kendon and Ferber, 1973).

### 3.3 Robot gestures to express mental states

In the first two studies the robot used non-verbal behaviors that were modeled on human behavior. We now describe a pilot study, conducted in the wild, that moves toward metaphorical gestures. This study focused on gestures to express the following mental states of the robot: *agreement*, *disagreement*, *uncertainty*, and *excitement*. In humans, facial expressions and head gestures play an important role in expressing these mental states. While SARAH can produce different facial displays, she does not have a movable head. Based on our intuition, we devised the following non-verbal behaviors.

*agreement* Smile and move forward and backward a few inches.

*disagreement* Frown and rotate side to side by 35 degrees.

*uncertainty* With a neutral facial expression, turn away from the addressee by 45 degrees, briefly pause, then return.

*excitement* With surprised facial expression, spin around 360 degrees.

SARAH recruited subjects in a busy hallway on campus. She instructed subjects to retrieve an index card with a set of yes/no questions from a pocket attached to the robot and to ask those questions. SARAH accompanied her spoken answer either with facial expressions only or with facial expressions and gestures. At the end of the scripted interaction, SARAH said "Yay, we completed the task" and expressed excitement.

SARAH then asked the subjects to complete a paper survey rating SARAH's intelligence and naturalness. In this pilot study, SARAH's use of gestures did not have a (statistically significant) impact on people's perceptions of her. And, unfortunately, we did not collect data that allows us to draw conclusions on whether humans interpreted the gestures as intended.

Interesting research problems that arise are the design of easy to interpret metaphorical gestures, how to select which signals to use in a given dialog situation, how to coordinate different communicative signals, and how to transition between and blend different non-verbal behaviors.

*Future work:* We are preparing a follow up study that will evaluate the interpretability of variants of different gestures more systematically. Our goal is to create a lexicon of robot behaviors that can perform different discourse and dialog functions. We are currently focusing on robot movements, but we are also interested in other signals, like non-speech sounds and visual cues on the screen that go beyond facial expressions mimicking humans.

## 4 Conclusion

The interactions between humans and robots in public spaces are situated in an un-controllable and only partially predictable environment. This creates challenges for communication. We think that NLG can contribute to a solution to these challenges by producing utterances and other communicative behaviors that are adapted to the situation. In addition, we argue that these challenges give rise to research problems that are interesting from an NLG point of view.

In this paper, we have illustrated our position by describing three studies that explore the generation of co-verbal communicative behaviors for non-humanoid robots. This line of research tackles the following issues related to the generation of multimodal utterances. We need to design non-verbal signals that are mimicking human behavior as well as signals that communicate metaphorically. The robot behaviors are constrained by the limited motor capabilities of the robot, but they can also take advantage of expressive options that are not available to humans. We need techniques for generating multimodal utterances that coordinate the different non-verbal signals and speech. And finally, we need to understand how to choose the most effective set of signals in a given dialog situation.

## References


James Allen, George Ferguson, and Amanda Stent. 2001. An architecture for more realistic conversational systems. In *Proc. of the 6th International Conference on Intelligent User Interfaces*.

Cynthia Lynn Breazeal. 2000. *Sociable machines: Expressive social exchange between humans and robots*. Ph.D. thesis, MIT.

Hendrik Buschmeier, Timo Baumann, Benjamin Dosch, Stefan Kopp, and David Schlangen. 2012. Combining incremental language generation and incremental speech synthesis for adaptive information presentation. In *Proc. of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*.

Aaron G. Cass, Eric Rose, Kristina Striegnitz, and Nick Webb. 2015. Determining appropriate first contact distance: trade-offs in human-robot interaction experiment design. In *Proc. of the Workshop on Designing and Evaluating Social Robots for Public Settings at the Intl. Conf. on Intelligent Robots and Systems*.

Aaron G. Cass, Kristina Striegnitz, Nick Webb, and Venus Yu. 2018. Exposing real-world challenges using HRI in the wild. In *Proc. of the 4th Workshop on Public Space Human-Robot Interaction at the Intl. Conf. on Human-Computer Interaction with Mobile Devices and Services*.

Justine Cassell, Joseph Sullivan, Elizabeth Churchill, and Scott Prevost. 2000. *Embodied Conversational Agents*. MIT press.

Elizabeth Cha, Yunkyung Kim, Terrence Fong, and Maja J. Matarić. 2018. A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends in Robotics*, 6(4):211–323.

Elizabeth Cha and Maja Matarić. 2016. Using nonverbal signals to request help during human-robot collaboration. In *Proc. of the Intl. Conf. on Intelligent Robots and Systems*.

Nicole Chovil. 1991. Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, 25(1-4):163–194.

Herbert H. Clark and Meredyth A. Krych. 2004. Speaking while monitoring addressees for understanding. *J. of Memory and Language*, 50(1):62–81.

Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and predictability of robot motion. In *Proc. of the 8th Intl. Conf. on Human-Robot Interaction*.

Paul Ekman and Wallace V. Friesen. 2003. *Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions*. Malor Books.

Joy E. Hanna and Susan E. Brennan. 2007. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *J. of Memory and Language*, 57(4):596–615.

Dai Hasegawa, Justine Cassell, and Kenji Araki. 2010. The role of embodiment and perspective in direction-giving systems. In *Proc. of the AAAI fall symposium: Dialog with robots*.

Kristiina Jokinen and Graham Wilcock. 2003. Adaptivity and response generation in a spoken dialogue system. In Jan van Kuppevelt and Ronnie W. Smith, editors, *Current and new directions in discourse and dialogue*. Springer.

Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.

Adam Kendon and Andrew Ferber. 1973. A description of some human greetings. In *Comparative Ecology and Behaviour of Primates: Proc. of a Conf. held at the Zoological Society London*. Academic Press.

Alexander Koller, Kristina Striegnitz, Donna Byron, Justine Cassell, Robert Dale, Johanna Moore, and Jon Oberlander. 2010. The first challenge on generating instructions in virtual environments. In E. Krahmer and M. Theune, editors, *Empirical Methods in Natural Language Generation*, volume 5980 of *LNCS*. Springer.

David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Verena Rieser, Oliver Lemon, and Simon Keizer. 2014. Natural language generation as incremental planning under uncertainty: adaptive information presentation for statistical dialogue systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(5).

Maha Salem, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing, and Frank Joublin. 2012. Generation and evaluation of communicative robot gesture. *Intl. J. of Social Robotics*, 4(2):201–217.

Cameron Smith, Nigel Crook, Simon Dobnik, Daniel Charlton, Johan Boye, Stephen Pulman, Raul Santos De La Camara, Markku Turunen, David Benyon, Jay Bradley, et al. 2011. Interaction strategies for an affective conversational agent. *Presence: Teleoperators and Virtual Environments*, 20(5):395–411.

Marilyn A. Walker, Amanda Stent, François Mairesse, and Rashmi Prasad. 2007. Individual and domain adaptation in sentence planning for dialogue. *J. of Artificial Intelligence Research*, 30.