Perception Based Misunderstandings in Human-Computer Dialogues

Niels Schütte
Dublin Institute of Technology
niels.schuette@student.dit.ie

John Kelleher
Dublin Institute of Technology
john.d.kelleher@dit.ie

Brian Mac Namee
Dublin Institute of Technology
brian.macnamee@dit.ie

Abstract

In a situated dialogue, misunderstandings may arise if the participants perceive or interpret the environment in different ways. In human-computer dialogue this may be due the sensor errors. We present an experiment system and a series of experiments in which we investigate this problem.

1 Introduction

Computer systems that engage in natural language dialogue with human users are known as **dialogue systems**. A dialogue system that operates in a spatial environment, a **situated dialogue system**, needs to have information about the spatio-temporal context. This can be achieved through perception of the environment. Perception, e.g. computer vision, always has the potential of producing errors, e.g. by failing to notice an object or by misrecognizing an object as a different type of object. We are interested in the effect that such perception-based errors have on human-computer dialogue. If the human user and the system have a shared view of the environment, false perception by the system will lead to a divergence between the user's understanding of the environment and the system's understanding and this in turn leads to problems in the interaction between the system and the user. For example, if the user asks a robot to pick up an apple, and the robot has mistaken a pear for an apple, it may instead pick up the pear. Misunderstandings of this kind also occur in human-human interaction and human speakers are able to establish and recover a shared understanding or common ground (Clark and Schaefer, 1989). Misunderstandings in human-computer dialogue due to misunderstandings because of problems in natural language understanding and speech recognition have been also been investigated and addressed (e.g. (Shin et al., 2002; López-Cózar et al., 2010)).

In an earlier work we investigated the problem of perception based misunderstandings in a corpus of data from human-human interaction (Schütte et al., 2012). In this paper we report on a work in progress in which we investigate the effect of sensor errors on human-computer dialogue using a dialogue system. Participants interact with a simulated robot through a text based dialogue interface in order to re-arrange objects in a virtual world.

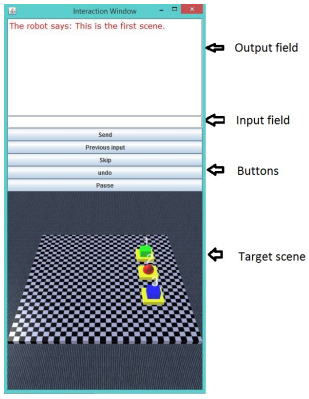
Participants are presented with a number of **scenes**. In each scene the participants are asked to instruct the robot arrange the objects present in the world into a given **target configuration**. Participants were given the option to abandon a scene if they felt they would not be able to complete it.

We perform a series of experiments that focus on three issues: (a) establishing a **baseline** for the difficulty of the interaction task (b) establishing the **impact of perception errors** on the baseline task performance and (c) establishing the **usefulness of different approaches to resolve the misunderstandings**.

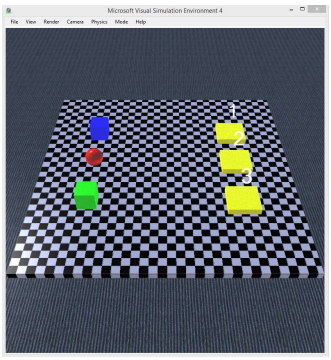
2 Experiment System

The experiment system consists of a dialogue system and a robot simulation environment. The dialogue system was implemented for this experiment and is focused on covering a wide range of spatial instructions. The robot is a highly simplified abstraction of a manipulator arm that can pick up objects and move them to specified locations. It is not rendered in the simulation.

Figure 1 shows the user interface presented to the participants. The left side shows the text based dialogue interface window. The image in the lower part of the window shows the scene the participant was asked to create. The right hand side of the figure shows the participants' view of the simulated world. In summary, the participants' task was to



(a) The dialogue interface.



(b) The simulation view.

Figure 1: The user interface.

interact with the system to change the scene in the window in Figure 1b into the scene displayed at the bottom of the window in Figure 1a.

The robot's perception of the world is provided through an abstract simulated vision system. By manipulating the vision system, targeted errors can be introduced into the system's perception. For example, it can be specified that the system mistakes the colour of a certain object for a different colour. If the participant now uses the colour to describe the object, the system will not be able to resolve the reference correctly. It should be noted that with this experiment we do not aim to produce a novel dialogue system or to provide an accurate simulation of computer vision, but to examine the performance of the the given system under different conditions of perceptual problems.

During each interaction, the contributions by the participant and the system are logged and annotated with their semantic interpretations. Parameters related to the dialogue such as task completion rate, number of actions, number of errors and time taken for each action are recorded. They serve as the basis of our comparison of the task difficulty of the different experiment conditions.

We are currently performing a series of three experiments with this experiment setup. **Experiment 1** serves to establish a baseline difficulty. It uses a series of scenes that were manually designed to encourage specific expressions. In **Experiment 2** errors are introduced into the robot's perception. This experiment serves to establish the impact of the errors on the interaction. Errors were manually designed for each scene to produce specific problem situations. In **Experiment 3** we evaluate different approaches towards solving the perception based misunderstandings by communicating the system's understanding of the scene to the user. The experiment uses the same scenes and errors as the second experiment. Participants are split into two groups. The first group is given the option of asking the system to describe verbally what it perceives. The second group is given the option of asking the system to visually communicate its understanding through the use of markup on the screen. Thereby both groups are given access to the system's understanding of the scene, but through different modalities.

3 Current State

We have currently finished the first two experiments and are evaluating the results. A first preliminary analysis and a more detailed description of the experiment will be available in (Schütte et al., 2014). The third experiment is currently commencing. A comparison of the results from the first and the second experiment indicates that the introduction of perception errors increased the difficulty of the task. Participants were much more likely to abandon scenes containing errors than scenes not containing errors. They also needed more actions to complete scenes with errors than scenes without errors, and often used more time doing so.

4 Future Work

After the completion of the third experiment, we are going to compare the results between the different experiments. We are planning to investigate the strategies used by the participants when they encountered problems in the dialogue and relate them to our work in human-human interaction.

References

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, pages 259–294.

Changsong Liu, Rui Fang, and Joyce Y. Chai. 2012. Towards mediating shared perceptual basis in situated dialogue. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, page 140–149. Association for Computational Linguistics.

Ramón López-Cózar, Zoraida Callejas, and David Griol. 2010. Using knowledge of misunderstandings to increase the robustness of spoken dialogue systems. *Knowledge-Based Systems*, 23(5):471–485, July.

Niels Schütte, John Kelleher, and Brian Mac Namee. 2012. A corpus based dialogue model for grounding in situated dialogue. In *Proceedings of the 1st Workshop on Machine Learning for Interactive Systems: Bridging the Gap Between Language, Motor Control and Vision (MLIS-2012).*, Montpellier, France, August.

Niels Schütte, John Kelleher, and Brian Mac Namee. 2014. The effect of sensor errors in situated human-computer dialogue. In *Proceedings of the The 3rd Workshop on Vision and Language (VL'14).*, page to appear, Dublin, Ireland, August.

Jongho Shin, Shrikanth S. Narayanan, Laurie Gerber, Abe Kazemzadeh, Dani Byrd, and others. 2002. Analysis of user behavior under error conditions in spoken dialogs. In *INTERSPEECH*.