Talking Robots: a Fully Autonomous Implementation of the Talking Heads

Jean-Christophe Baillie & Matthieu Nottale
Laboratory of Electrical and Computer Engineering, ENSTA
32 Bd Victor 75015 Paris France
jean-christophe.baillie@ensta.fr   matthieu.nottale@ensta.fr

Abstract

The “Talking Robots” experiment, inspired by the “Talking Heads” experiment from Sony, explores possibilities on how to ground symbols into perception using language, with two autonomous Aibo robots in an unconstrained environment. We present here the first results of this experiment and outline in the conclusion a planned extension to social behaviors grounding.

1 Introduction

The “Talking Robots” experiment, inspired by the “Talking Heads” experiment from Sony, has been started for six months in our laboratory. The purpose of this experiment is double: first, to show that the original “Talking Heads” experiment can be reproduced in a more demanding context: autonomous robots (Aibos), unconstrained vision and speech, and autonomous synchronization. Second, to serve as a ground to build a more complex experiment. Making a full use of the unconstrained context, we will in future work investigate how the robots could autonomously develop their own language games.

We will present briefly in this poster the structure of the existing “Talking Robots” experiment and the underlying scientific questions that we address. We will also comment on our first results.

2 Grounding symbols into perception

Grounding symbolic representations into perception is a key and difficult issue for artificial intelligence [Harnad, 1990; Steels and Baillie, 2003]. Including social interactions and, more specifically, language acquisition and development in this context has proven to be a fruitful orientation. One of the recent and successful attempts in this direction is the ”Talking Heads” experiment [Steels and Kaplan, 2002]. This experiment involves two cameras interacting in a simplified visual environment made of colored shapes on a white board. The agents inside the cameras are developing a *shared grounded lexicon* of words related to visual meanings like ”red”, ”large”, ”above”, etc.

The ”Talking Heads” experiment has proven that it was possible to design a computer-based mechanism for language acquisition that shares many properties in common with what can be observed in human language acquisition. It can be seen as an attempt in the direction of a computational model of language acquisition and symbol grounding, outlining possible key structural elements that might be essential for such a development.

However, the ”Talking Heads” had two main limitations, which were good and necessary simplifications to start with, but that we will consider carefully here:

- The environment was limited and simplified (only colored shapes on a white background, fixed cameras).
- The interaction protocol (language game) was predetermined and hardcoded.

With the recent development of relatively cheap and powerful robotic platforms (see Sony, Honda or Fujitsu, among others) research on symbol grounding can move from simulation or simple environments to complex natural environments and embodied systems. Following this trend, we have started the ”Talking Robots” experiment which reproduces the initial Talking Heads experiment with autonomous robots (Aibo ERS7) evolving in an unconstrained visual environment instead of simple cameras.

The second point, regarding hardcoded interactions, is not handled in the Talking Robots, but constitutes our future research goal. The visual routines, robust perception mechanisms and the general experience we drew from the Talking Robots will be a key component and a basis for an extended experiment involving the dynamical creation of language games between two robots.

3 Results

3.1 Technical issues

The autonomous and unconstrained implementation of the language game called ”Guessing Game” involves the following technical and practical issues:

- Automatically determine the leader (the ”speaker”) and the follower (the ”hearer”) in the game: we use beep signals with specific tones to perform synchronization.
- Have the two robots standing one next to the other, starting from a random position: we use movement detectors and a waving head motion to locate an Aibo head in the

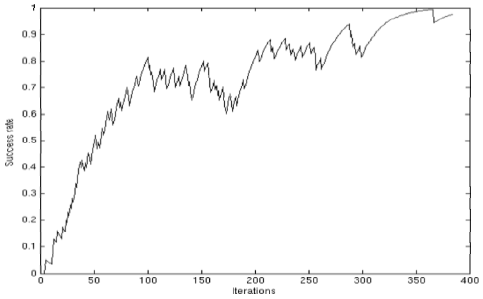
image and pattern recognition based on a FMI-SPOMF transform[sheng Chen *et al.*, 1994] to identify the orientation of this head. Together with a scale measure, this gives an accurate enough estimation of the position and orientation of the other robot.

- Use a stable segmentation algorithm which will give comparable results from one robot to the other, if they look in approximately the same direction: the CSC algorithm[Rehrmann and Priese, 1998] has proven to be the most stable after a series of tests (the measure of stability is the entropy of a normalized correlation matrix between two segmentations of two slightly different images, which should ideally be zero).
- Provide an accurate pointing device for the robots to designate the objects they are talking about: we use a blinking laser pointer fixed on the robot's head and a simple red spot detector which gives excellent results.

Putting all these elements together, we successfully obtained converging results for the "Discrimination Game" (comparable to the original experiment) and we have run several successful "Guessing Games".

3.2 Discrimination Game

The discrimination game[Steels and Kaplan, 2002] is played by one robot only and is designed to dynamically create a set of visual categories grounded in the surrounding visual environment of the robot: for example, with objects of different sizes in the image, a category for "big" and "small" will be developed but no category corresponding to "colors" will arise in a black and white environment. Our games successfully converge with 250 iterations to a set of approximatively 80 to 90 categories in the unconstrained environment of the lab, and a success rate above 80%:

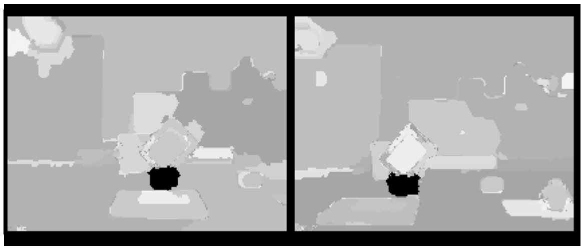


Categories are mainly developed for position and area. If we turn the vision of the robot to black and white, the discrimination tree corresponding to "saturation" shrinks in a few iterations, as expected.

3.3 Guessing Game

The "Guessing Game" has been run on a limited set of experiments. Although we have still no statistical results yet, we have observed in detail several games to compare them to the original Talking Heads.

The following figure shows the result of a guessing game played on a topic (in black) whose meaning has been misunderstood by the hearer but correctly identified after the speaker has pointed it (both robots succeed in sharing the same topic):



This kind of failure is currently happening most of the time, and we have to improve the constancy of segmentation and context to increase the probability of having the speaker and the hearer both categorizing a topic in the same way. Apart from this, the original mechanisms of the Talking Heads have been successfully reimplemented and run. Further results will be presented in the poster.

4 Conclusion

The Talking Robot experiment has started to successfully reproduce in a complex environment the results obtained in the simplified environment of the Talking Heads. Building on our experience on this embedded version, we plan to start a new set of experiments involving not only a predetermined *language game* interaction protocol, but where the robots will dynamically create a protocol and use it to play a language game which will not be known in advance. This ambitious research project will require to state precisely the core structural elements of the Talking Robots and propose an architecture to generalize them and have a mechanism to let them evolve. On the practical side, we will make use of the robust vision and localization algorithms we developed for the Talking Robots.

References

[Harnad, 1990] Harnad. The symbol grounding problem. *Physica D 42*, pages 335–346, 1990.

[Rehrmann and Priese, 1998] Rehrmann and Priese. Fast and robust segmentation of natural color scenes. In *ACCV (1)*, pages 598–606, 1998.

[sheng Chen *et al.*, 1994] Qin sheng Chen, Michel Defrise, and F. Deconinck. Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(12), 1994.

[Steels and Baillie, 2003] Luc Steels and Jean-Christophe Baillie. Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems*, 43(2-3):163–173, 2003.

[Steels and Kaplan, 2002] Steels and Kaplan. Bootstrapping grounded word semantics. In T. Briscoe, editor, *Linguistic evolution through language acquisition: formal and computational models*, chapter 3, pages 53–73. Cambridge University Press, Cambridge, 2002.