Video links:
https://youtu.be/FpYDco3ZgkU
http://ijcai-15.org/downloads/videos/competition/11-Humanoidly-Speaking.mp4

Accepted video of the International Joint Conference on Artificial Neural Networks – IJCAI, Video Competition, 2015.
Title: Humanoidly Speaking – How the Nao humanoid robot can learn the name of objects and interact with them through common speech
Authors: Xavier Hinaut, Johannes Twiefel, Marcelo Borghetti Soares, Pablo Barros, Luiza Mici, Stefan Wermter

University of Hamburg, Department of Computer Science, Knowledge Technology, www.knowledge-technology.info

The video is based on the research of the projects EchoRob, CoRob, and DOCKS.

Abstract:
This video shows a friendly human-robot interaction using humanoid Nao robots. The speaker teaches the robot some names of objects using speech. This work shows the successful integration of three different projects mainly using Artificial Neural Networks: (1) object recognition with RGB-D (color and depth) sensor, (2) speech to text using an approach that post-processes Google's speech recognition hypotheses, and (3) syntactic interpretation of sentences.

The robot is able to identify surfaces in the environment (tables, floor, walls) and establish a relation between these surfaces and the clusters (objects). Multiple viewpoints are easily obtained from the segmented clusters and used for training a Convolutional Neural Network. The features obtained allow the robot to recognise objects and to generalise to unknown viewpoints and scales.

The speech recognition system maps the results from Google to expectable sentences in the given scenario using phonemic matching. The syntactic interpretation of the sentence is done with a Recurrent Neural Network (namely an Echo State Network). It maps each semantic word in a sentence to its thematic role. In the end, all roles form predicates which indicate what should be performed (e.g. learning a new object or performing motor actions).

At the start, the robot does not know any objects. During the learning of new objects, increasingly complex sentences are used to describe the position of new objects. Motor commands (e.g. pointing) are also provided in order to check the knowledge of the robot. It can be noted that the human user produces natural complex sentences, and thus any human could interact with the robot, not only robot programmers. Furthermore, complex sentences containing multiple commands can be correctly interpreted as a temporal action sequence (e.g. "Before doing 'B' do 'A'") without adding any complementary mechanism.

In addition to teaching objects and relationships to the robot, in the future this kind of interaction scheme could also be used by children to learn (e.g. new objects) while interacting with the robot. In other words, by teaching the robot, the child is learning. It could be used also to teach new instructions or even new languages to the robot (with only few changes in

modules (2) and (3)). Conversely, if the robot already knows about the environment in one language and a child would not, this child could learn this new language while interacting with the robot.

Acknowledgments:
This research was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme: EchoRob project (PIEF-GA-2013-627156).