Summary for the Final Report for EchoRob

The EchoRob project has led to four major outcomes.

First, we were able to demonstrate that a biologically plausible model of sentence comprehension based on recurrent neural networks, called reservoirs, could (1) manage to learn and generalize (to about 80% of unknown sentences) on two languages at the same time (English and French) by using the very same random RNN instance: To our knowledge, this is the first time that the same instance of such a fixed random neural network based model was able to learn basic syntax in different languages with a linear read-out. This is especially interesting from a child language acquisition perspective, showing that no pre-wired structure is necessary to acquire different grammars. Moreover, (2) the model could deal with Out-Of-Vocabulary words on learned sentences: when implemented in a robot, this ability permits to still process sentences correctly, even if the speech recognition pre-processing did not recognize a word.

Secondly, we proposed an efficient reservoir-based parser able to generate a single semantic tree structure in real-time, which can be executed by a robot arm manipulating objects. The system outputs a structure of action events enabling it to represent complex sentences. We outperformed four of six other approaches, with 74.1% accuracy under best conditions, on a noisy and ambiguous corpus of robotic commands generated by online users, which demonstrated that the model is able to cope with realistic noisy conditions.

As a third outcome that will specifically help future work, we designed the structure of a crowdsourcing website in order to collect Human-Robot Interaction corpora from online users with a particular focus on user motivation over time to increase the performance for the tasks leading to high quality data. A beta version of the website was implemented as a local prototype with which we started to successfully collect natural sentences from users in many various languages such as Arabic, Romanian, Farsi, Vietnamese or Urdu.

A final outcome, which also had the highest impact during the project duration, was a multi-modal human-robot interaction architecture conjointly developed and implemented on the Nao humanoid robot. It combines the neural language and convolutional neural networks to integrate 3D vision, speech and sentence interpretation. It is able to generate feedback to the user which helps to implicitly understand how to interact with the robot. During experiments with subjects, the speech module achieved an accuracy of 92.4% together with the language module. The system was also able to detect inconsistent input coming from different sensory modules in all cases and could generate useful feedback for the user. The overall architecture was shown to the public on the Night of Knowledge 2015 in Hamburg, where people interacted with the Nao over several hours without interruption, demonstrating the robustness of the system in a very noisy environment of hundreds of people. This idea of sentence comprehension with a reservoir-based model was also extended to 3D-body joint action recognition, showing the versatility of the approach.

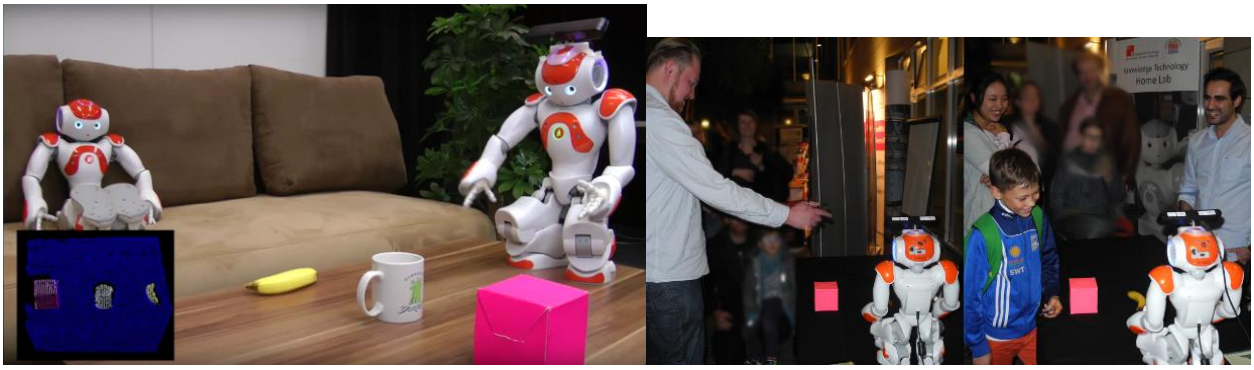
Overall, this project demonstrated that reservoir-based neural networks applied to language processing could be of use in multiple domains: Such a model (1) can give insights on how sentences are processed and produced by our brains, which leads to a better comprehension of the underlying mechanisms, and

could lead to advances in rehabilitation and better care of people with language brain lesions, for instance; (2) provides a better understanding of how children acquire languages, which is a crucial question in psycholinguistics (e.g. the Chomskian nativist debate) and which could also help to improve learning methods for language learning; (3) could provide a universal way of processing multiple languages with the same system and thus provide services to some languages poorly translated across the world; (4) will, on the computer science and robotics side, provide new ways to process speech and language in real-time with low consumption of resources: reservoir methods do not need remote super-computers to work, they could work online on on-board chips (e.g. on a robot architecture) in a robust way; Finally, (5) it could provide a new way of designing architectures for Human-Robot Interaction, for instance providing the ability to people to customize their robots or smart-objects by adapting them in a natural way to their particular way of speaking (which could be very different if we consider for instance the British, American or Indian way of pronouncing English). Similar architectures could be used for other modalities than language such as human gesture.

EchoRob project: https://www.informatik.uni-hamburg.de/WTM/projects/EchoRob/EchoRob.shtml

Explanatory video of the implementation in Nao robot: https://youtu.be/FpYDco3ZgkU

EchoRob Logo and image project.



Nao robot architecture based on the neural language model: (left) Demo and explanatory video, (right) Demo to the public at the Night of Knowledge 2015 in Hamburg.