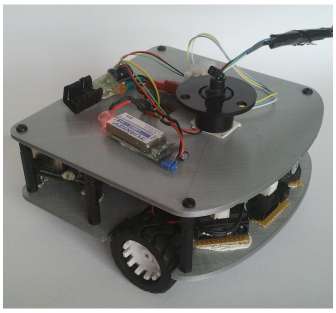
Learning and Imitating Robot Ego-Noise

Antonio Pico*, Guido Schillaci*, Verena V. Hafner* and Bruno Lara**

(*) Adaptive Systems Group, Humboldt-Universität zu Berlin, Germany.
E-mails: {pivilla, guido.schillaci, hafner}@informatik.hu-berlin.de
(**) Cognitive Robotics Group, Universidad Autónoma del Estado de Morelos, Mexico.
E-mail: bruno.lara@uaem.mx



In this poster, we present a mechanism for learning, predicting and imitating ego-noise on a custom robotic platform that we designed and built in our department. Robot ego-noise is the noise that the robot itself generates while moving around. We adopted forward and inverse models [1], [2] as a computational tool for:
1- encoding the dynamics of the motor system of the robot,
2- predicting the effect of self-produced movements on the perceived noise distinguishing noise and ego-noise (forward models) and,
3- inferring the motor commands needed for reproducing a specific ego-noise (inverse models).

The models were trained using sensorimotor data gathered by executing a self-exploration behaviour on the robot, namely random motor babbling. Data consisted of proprioceptive and exteroceptive information, in addition to motor information. The aim of the training is to learn the mapping between these different modalities. Proprioceptive data was gathered from speed sensors located on each wheel of the robot. At time *t*, a sensory state *S(t)* is encoded by two variables (sensed speed from the left and right wheels). In the experiments presented here, the inverse model was fed with three sensory states as inputs: *S(t), S(t-1) and S(t-2)* , that is with a vector of 6 elements. Exteroceptive data consisted of the auditory response *S(t+1)* of a motor command M(t) applied from the initial state *S(t)*. The Auditory sensory situation *S(t+1)* was formed as a vector of 12 Mel-frequency Cepstral Coefficients [3] extracted from an audio chunk of 100ms, recorded from time *t* to time *t+1*. Motor information encodes the motor command applied to each wheel so that at time *t*, *M(t)* is represented by two variables. The output of the inverse models are then two motor commands (a 2 element vector).

The inverse model was implemented as a motor map, an extension of Kohonen's Self-Organizing Maps (SOM). This motor map consists of a SOM of sensory inputs matched with a SOM of motor outputs in such a way that each input gets associated to the correct motor output [4]. The motor map was trained with 20000 samples of collected data.

We performed a preliminary experiment on imitating robot ego-noise, which consisted in making the robot execute a series of predefined motor commands while gathering audio data. Then, the robot had to reproduce the recorded ego-noise by applying the motor commands generated by the inverse model, which was fed with the speed of the motors and the previously gathered audio data.
We are currently performing additional experiments and gathering quantitative results in order to demonstrate the imitative capabilities of the proposed system.

References:

[1] D. M. Wolpert, Z. Ghahramani, and J. R. Flanagan, "Perspectives and problems in motor learning," Trends in Cognitive Sciences, vol. 5, no. 11, pp. 487 – 494, 2001.
[2] G. Schillaci, V. V. Hafner, and B. Lara, "Exploration behaviours, body representations and simulations processes for the development of cognition in artificial agents." Frontiers in Robotics and AI, vol. 3, no. 39, 2016.
[3] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," Journal of Computing, vol. 2, no. 3, pp. 138–143, 2010.
[4] Ritter, Helge J., and Klaus Schulten. "Topology-conserving maps for motor control." *Neural Networks* 1.Supplement-1 (1988): 357.