Preprint. Accepted as a conference paper at the *International Conference on Intelligent Robots and Systems* (IROS) 2018, ©IEEE 2018

Hear the Egg — Demonstrating Robotic Interactive Auditory Perception

Erik Strahl[1], Matthias Kerzel[1], Manfred Eppe[1], Sascha Griffiths[1] and Stefan Wermter[1]

Abstract

We present an illustrative example of an interactive auditory perception approach performed by a humanoid robot called NICO, the Neuro Inspired COmpanion [1]. The video demonstrates a material classification task in the style of a classic TV game show. NICO and another candidate are supposed to determine the content of small plastic capsules that are visually indistinguishable. Shaking the capsules produces audio signals that range from rattling stones, over tinkling coins to swooshing sand. NICO can perceive and analyze these sounds to determine the material of the capsule's content.

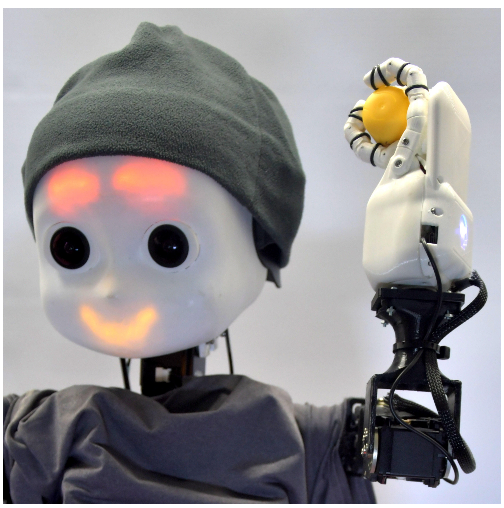

Fig. 1. THE NICO robot actively shaking a yellow plastic capsule to distinguish the content from the recorded audio

Though presented in an edutainment context, the ability to distinguish objects and their content by eliciting and classifying audio signals is not only helpful but in fact necessary for a humanoid companion robot that assists humans in an unstructured domestic environment. By using interactive auditory perception [2], a robot can detect if a salt shaker is empty or the box of biscuits really contains cookies without opening it for visual inspection.

In an article [3] which we have submitted to this conference as a regular paper, we describe how a novel recurrent neural network approach is developed to learn and perform the auditory classification from recorded sound information. The sound is generated by the NICO robot in an interactive perception process by shaking the yellow, egg-shaped, capsules (see Figure 1). These capsules have been filled with the ten different materials depicted in Figure 2.

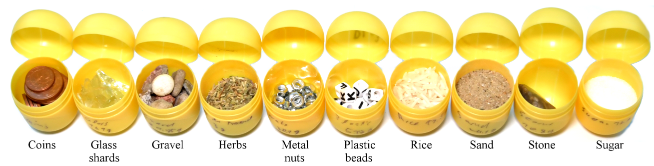

Fig. 2. The yellow egg-shaped capsules contain ten different materials

While shaking the capsules, the NICO robot records a 625 milliseconds chunk of audio with the microphones integrated into its ears. For our dataset we recorded 1080 of such signal samples, using 1000 signals to train a neural network and 80 signals to evaluate it. The audio was preprocessed using Mel Frequency Cepstral Coefficients (MFCC) before passing it to a recurrent neural network. For the demo, we used Long Short-Term memory (LSTM) cells but found later that Gated Recurrent Units (GRU) perform even slightly better, achieving an overall average classification performance of 91% (see Figure 3).

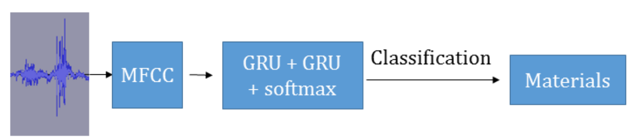

Fig. 3. Architecture used for the material classification

[1]Knowledge Technology, Department of Informatics, University of Hamburg, Germany. {strahl, kerzel, eppe, griffiths, wermter} @informatik.uni-hamburg.de

The authors gratefully acknowledge partial support from the German Research Foundation DFG under project CML (TRR 169) and the Hamburg Landesforschungsförderungsprojekt CROSS.

References

[1] M. Kerzel, E. Strahl, S. Magg, N. Navarro-Guerrero, S. Heinrich, and S. Wermter, "Nico - neuro-inspired companion: A developmental humanoid robot platform for multimodal interaction," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2017, pp. 113–120.

[2] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive Perception: Leveraging Action in Perception and Perception in Action," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1273–1291, dec 2017.

[3] M. Eppe, M. Kerzel, E. Strahl, and S. Wermter, "Deep Neural Object Analysis by Interactive Auditory Exploration with a Humanoid Robot," in *International Conference on Intelligent Robots and Systems (IROS) – under submission, accepted*, 2018.