



SCIENCES &  
BIOENGINEERING SCIENCES

The Research Group

## Artificial Intelligence Lab (ARTI)

has the honor to invite you to the public defense of the joint PhD thesis  
of

**Heikki RASILO**

to obtain the degree of Doctor of Sciences

Joint PhD with Aalto University (Finland)

Title of the PhD thesis:

*Computational models of associative learning in the acquisition of speech imitation, acoustic word models and word-meaning mappings*

Promotors:

Prof. Bart de Boer  
Prof. Paavo Alku (Aalto University)

The defence will take place on

November 3 2017 at the Aalto University,  
Finland

Members of the jury:

Dr. Okko Räsänen (chairman, Aalto)  
Prof. Bart de Boer (secretary)  
Prof. Werner Verhelst  
Prof. Katrien Beuls  
Prof. Unto K. Laine (Aalto)  
Prof. Dr. Bernd Kröger (RWTH Aachen)  
Prof. Louis ten Bosch (Univ. Nijmegen)

### Curriculum vitae

Master of Science in Technology,  
Aalto University School of Science  
and Technology, Faculty of  
Electronics, Communications and  
Automation 2004 - 2010.

A doctoral student at Aalto  
University, Department of Signal  
Processing and acoustics and the  
VUB 2010 –2017.

### Abstract of the PhD research

Human infants manage to learn their native language from noisy and ambiguous language input, but the exact mechanisms underlying language learning are not known. Recent decades of research have shown that so-called statistical learning mechanisms provide a possible way to learn linguistic patterns from sensory signals, either within auditory input or as consistent relations between different perceptual domains. Statistical learning is especially powerful in situations where relevant patterns in different sensory streams do not occur in a one-to-one fashion, enabling the learning of associations that still occur at above-chance level. In this thesis, the potential role of associative statistical learning is studied in the context of three aspects of language learning: learning to imitate the speech of human caregivers, learning words and their segmentation from continuous speech, and learning word-to-meaning mappings. The three tasks are investigated with computational models of human-like learning in ambiguous learning situations.

The first learning task considers the existing hypothesis that infants have to learn to imitate the speech of their parents, rather than the imitation skill being innate. It is hypothesized that infants can use statistical relations between their vocalic babble and caregivers' imitative responses to learn correspondences between the two. In this thesis, a mechanism for the learning of vocal imitation with ambiguous babble-response type input is introduced, and tested using human participants acting as caregivers to a virtual infant. It is also investigated how ambiguous visual information about possible word meanings can be used to bootstrap the learning of acoustic word models and word segmentation in continuous speech. Finally, it is studied what kind of cognitive constraints could explain human behaviour in so-called cross-situational learning experiments, where the subjects need to infer correct word-to-meaning mappings from ambiguous pairings of audiovisual stimuli.

The findings of this thesis indicate that statistical associative learning can be successfully used in several tasks related to language learning, and that highly specialized innate mechanisms related to speech may not be necessary in order for speech learning to take place. The findings and the computational algorithms introduced in this thesis may be of technological use in the implementation of autonomous robots able to learn from their environment, while also offer insight into what learning mechanisms may exist in the human brain and what kinds of stimuli are beneficial in order to facilitate human speech learning.