

Open Thesis / Project

Robust Speaker Identification using Resource-Constrained Embedded Devices

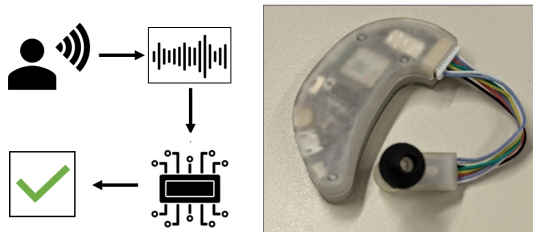
Thesis Type

Master Project / Master Thesis

Motivation

With the emergence of powerful speaker recognition systems such as Alexa and Siri, people are becoming fond of using voice activation instead of physical input. Together with the rapid growth of the Internet of Things (IoT), an increasing number of embedded devices are being deployed in various settings that could benefit from voice activation. However, embedded IoT devices are often not as powerful as Alexa or Siri, and only have a *fraction* of their memory and processing power.

Our goal is to let resource-constrained embedded IoT devices (i.e., deeply embedded devices with limited memory and computational resources) perform speaker recognition autonomously: this could be used to detect people's presence in a room (e.g., to easily clock in and out of work), or to eliminate the need for bulky keyboards or pin code inputs. Instead, only a small and cheap microphone is needed. Towards this goal, we have shrunk a state of the art model to fit the resource constraints of low-power IoT devices while retaining on-par accuracy. We are now interested on its deployment, on improving its robustness and usability, and on benchmarking its performance using different platforms, such as the OpenEarable 2 (a fully open-source AI platform for ear-based sensing applications) and the NPU-enabled Grove Vision AI Module V2.



Goals and Tasks

Within this context, students can explore several directions and perform different tasks, such as:

- Understand how state-of-the-art speaker recognition systems work, and how they can be shrunk to fit on embedded devices;
- Explore different enhancements to improve robustness and usability, for instance, training-data augmentation techniques, comprehensive benchmarking under varied acoustic conditions, or alternative model-training strategies;
- Leverage a second microphone aimed toward the ear to separate external noise from speech and perform active noise-cancellation, and other sensors to confirm that the wearer is speaking;
- Develop a robust speaker-recognition system prototype running on a next-generation IoT device equipped with an on-chip Neural Processing Unit (e.g., an integrated Arm Ethos-U55).

Target Group

- Students of ICE/Telematics;
- Students of Computer Science;
- Students of Electrical Engineering.

Required Prior Knowledge

- Basic knowledge of machine learning;
- Solid skills in Python and C programming;
- Experience with microcontrollers.

Contact Person

- Dr. Jesús Pestana Puerta
jesus.pestanapuerta@tugraz.at
- Assoc.Prof. Carlo Alberto Boano
cboano@tugraz.at



4480 – Institute of Technical Informatics (ITI)

Low-Power Embedded Networked Systems (LENS) Group
Group leader: Assoc.Prof. Carlo Alberto Boano



LENS