7a.026.TUT - Improving Speech Recognition Robustness

Project - Team

Team Member	Role	Email	Phone Number	Academic Sites/Industry Members
Moncef Gabbouj	PI	moncef.gabbouj@tuni.fi	+358 (400) 736613	Tampere University
Okko Rasanen	Co-PI	okko.rasanen@tut.fi	Not available	Tampere University
Mohammad Al-Sad	Student/Researcher	mohammad.al-sad@tut.fi	Not available	Tampere University
Ali Senhaji	Student/Researcher	Not available	Not available	Tampere University
Jaakko Vainio, Filip Ginter and Peter Sarlin	Project Mentors	Not available	Not available	**Funded By: Silo.ai**
				Funded By: Business Finland

Project - Summary

Speech recognition for general tasks is widely available in many languages by providers such as Google and Microsoft. However, current level technology has not been able to come up with a good quality general speech recognizer. Therefore, a good quality system requires training on case specific datasets for the task at hand. This is not possible with e.g. Google Speech API. Another aspect of this customization is the ease with which it is possible to take additional aspects of the audio into account on subsequent models. Another concern with these cloud services is confidentiality. In some use cases the data cannot be allowed to leave the organization in question. The costs of continued use of the cloud services can also be considerable. The aim of this project is to help Silo.ai to create an in-house solution for speech recognition. A general deep learning -based speech recognizer is trained on open data and other available sources. The general model is used in the creation of better case specific models, which are trained on client data. The resulting models are portable and can be setup in either cloud environments or local servers. The model can easily be combined with or serve as an input for additional ML models, such as sentence classification.

Robustness to noise and other interference in the environment is a crucial feature of a speech recognition system. The system should also be robust to different speakers, especially in public environments, instead of being adapted specifically to each user. The purpose of this research project is to investigate different approaches to make speech recognition systems robust to noisy environments and different speakers. This project will study speech enhancement techniques with next-generation, data-driven approaches.

Project - Novelty of Approach

Existing methods have studied speech recognition using several machine learning techniques, such as Generative Adversarial Networks. In this project, we wish to make these algorithms less prone to noisy environments and insure its robustness to different speakers (other than the user). To this end, we aim to develop advanced machine learning based on our recently developed GOP and POP neural networks.

Project - Deliverables

	Deliverable
1.	A new Database based on public domain resources for Automatic Speech Recognition task in Finnish language
2.	A baseline for the task and determine appropriate metrics for robustness
3.	Augment the merged database to expand the engine training domain and to test its ability in generalizing to unseen environments, speakers and samples

Table of Contents

4.	Propose a novel data augmentation technique to enhance the deep learning approach for Automatic Speech Recognition Systems in general and for Finnish specifically

Progress to Date

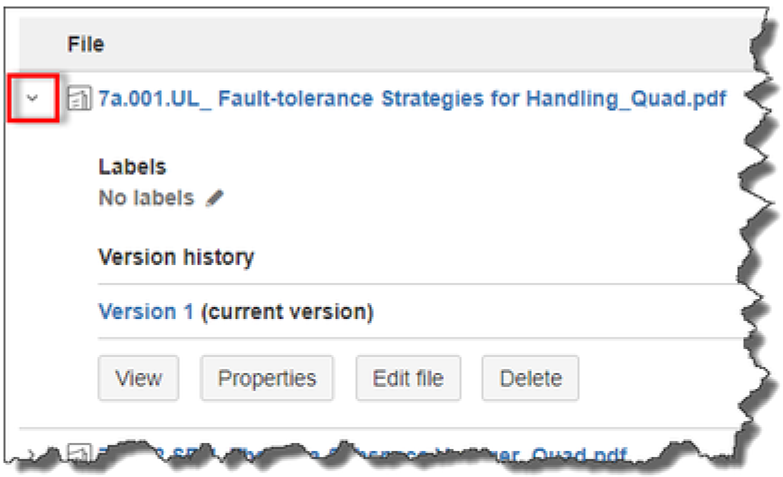
A rich set of machine learning algorithms has been developed by TUT in the past 5 years. Recently, we have tested existing multi-modal data fusion (both shallow and deep) methodologies and developed novel optimization criteria for extending performance in classification and retrieval applications.

Project - Benefits to IAB

The project can improve speech recognition in noisy environment and against other speakers present in the environment. As a result, human-machine interfaces will gradually replace text input relieving the user to attend to other more pertinent tasks.

Project - Documents

For viewing/editing options, please click left arrow next to document name.You will see different options depending on your access level.



File	Modified
7a.026-TUT-Improving-speech-recognition-robustness-Executive-Summary-modified-no-budget.docx	Oct 18, 2018 by Sally Johnson
7a.026-TAU-CVDI-Mid-Year-Report-ASR.pdf	Jan 23, 2019 by Sally Johnson

Drag and drop to upload or browse for files

Download All

CVDI Marketing Materials
CVDI Reports & Document Library
CVDI SITE (Drexel University)
CVDI SITE (Stony Brook University)
CVDI SITE (Tampere University)
CVDI SITE (University of Louisiana at Lafayette)
CVDI SITE (University of North Carolina at Charlotte)
CVDI SITE (University of Virginia)
IAB - Industry Advisory Board
Year 6 - Funded Projects (7/1/17 - 6/30/18)
Year 7 - Funded Projects (7/1/18 - 6/30/19)
Year 8 - Proposed Projects