DT2119 Speech and Speaker Recognition 7.5 credits

Igenkänning av tal och talare

This is a translation of the Swedish, legally binding, course syllabus.

Establishment

Course syllabus for DT2119 valid from Spring 2017

Grading scale

A, B, C, D, E, FX, F

Education cycle

Second cycle

Main field of study

Computer Science and Engineering

Specific prerequisites

For non-program students, 90 credits are required, of which 45 credits should be in mathematics or computer science. Furthermore, English B or the equivalent is required.

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

The student should, on completion of the course, be able to

- use the, in the course described, methods to recognise speech or speakers
- configure a system to a given application
- adapt and develop existing systems for speech and speaker recognition
- evaluate systems for speech and speaker recognition
- carry out research in the area.

Course contents

The course consists of lectures, three laboratory sessions with hand-in assignments, as well as writing a thesis in a subject chosen in consultation with the teacher. The thesis is furthermore presented orally during a final seminar. The laboratory sessions consist of designing different parts of a speech recognition application, train the system and evaluate its performance.

The following theoretical components are included:

- algorithms for training, recognition as well as adaptation to properties of speakers and transmissions channel, including pattern recognition, Hidden Markov Models (HMMs) and Deep Neural Networks (DNNs)
- methods to decrease the sensitivity against disturbances and deviations
- probability theory
- signal processing and parameter extraction
- acoustic modelling of the static and dynamic spectral properties of the speech sounds
- statistical modelling of language in spontaneous and formal speech
- search strategies- basic methods and strategies for large vocabularies
- specific analysis and decision making methods for recognition of speakers.

Furthermore, certain practical insight to build an application is given. Here, implementing certain functions based on prototypes and testing them on real speech data are included.

Course literature

- Huang, X., Acero, A., Hon, H.-W. Spoken Language Processing – A Guide to Theory, Algorithm and System Development, Prentice Hall, 2001.
- Automatic Speech Recognition: A deep learning approach, Dong Yu and Li Deng, Springer 2015. You can download the PDF through KTH Library.
- Särtryck på artiklar inom talarigenkänning och andra ämnen som inte ingår i boken.

Examination

- PRO1 - Project, 3.0 credits, grading scale: A, B, C, D, E, FX, F
- LAB1 - Computer Lab, 4.5 credits, grading scale: P, F

Based on recommendation from KTH’s coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

If the course is discontinued, students may request to be examined during the following two academic years.

Other requirements for final grade

Laboratory exercises Written assignments Thesis with presentation at a final review Assessment of two other course participants' theses and critical review on their presentations.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.