



FDT3119 Igenkänning av tal och talare 7,5 hp

Speech and Speaker Recognition

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

Fastställande

Kursplan för FDT3119 gäller från och med VT19

Betygsskala

P, F

Utbildningsnivå

Forskarnivå

Särskild behörighet

Doktorander från EECS

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Studenten ska efter genomgången kurs kunna:

- implementera träning och utvärderingsmetoder för taligenkänning

- träna och utvärdera en taligenkännare med hjälp av mjukvarupaket
- jämföra olika särdragsextraktioner och träningsmetoder
- dokumentera och diskutera specifika aspekter relaterade till tal- och talarigenkänning
- med hjälp av litteraturen, granska och kritisera andra studenters arbete i ämnet.

Kursinnehåll

Kursen består av föreläsningar, tre laborationer med inlämningsuppgifter, samt att skriva en uppsats i ett ämne valt i samråd med läraren. Uppsatsen presenteras dessutom muntligt under ett slutseminarium. Laborationerna består i att designa olika delar av en taligenkänningsapplikation, träna systemet och utvärdera dess prestanda.

Följande teoretiska delmoment ingår:

- algoritmer för träning, igenkänning samt adaptation till egenskaper hos talare och transmissionskanal, inklusive mönsterigenkänning, Hidden Markov Models (HMMs) och Deep Neural Networks (DNNs)
- metoder för att minska känsligheten för störningar och avvikelser
- sannolikhets teori
- signalbehandling och parameterextraktion
- akustisk modellering av talfjudens statistiska och tidsvarierande spektrala egenskaper
- statistisk modellering av språkbruk i spontant och formellt tal
- sökstrategier - grundläggande metoder och strategier för stora vokabulärer
- specifika analys- och beslutsmetoder för igenkänning av talare

Dessutom ges viss praktisk inblick i att bygga en tillämpning. Har ingår att implementera vissa funktioner utifrån prototyper och att testa dem på riktig taldata.

Kursupplägg

12 föreläsningar, 3 laborationer, projekt

Kurslitteratur

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.

Research articles in speech recognition

Examination

- EXA1 - Tentamen, 7,5 hp, betygsskala: P, F

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enstaka studenter.

LAB1 - Laboration, 4,5, betygsskala: P, F

PRO1 - Projekt, 3,0, betygsskala: P, F

Övriga krav för slutbetyg

Laboration med muntlig presentation

Forskningsprojekt med skriftlig rapport

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.
- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.