

EE-554

Automatic speech processing

Magimai Doss Mathew

Cursus	Sem.	Type
Computer science	MA1, MA3	Opt.
Cybersecurity	MA1, MA3	Opt.
Data Science	MA1, MA3	Opt.
Digital Humanities	MA1, MA3	Opt.
Electrical and Electronical Engineering	MA1, MA3	Opt.
SC master EPFL	MA1, MA3	Opt.

Language of teaching	English
Credits	4
Session	Winter
Semester	Fall
Exam	Written
Workload	120h
Weeks	14
Hours	4 weekly
Lecture	2 weekly
Exercises	2 weekly
Number of positions	

Summary

The goal of this course is to provide the students with the main formalisms, models and algorithms required for the implementation of advanced speech processing applications (involving, among others, speech coding, speech analysis/synthesis, and speech recognition, speaker recognition).

Content

1. Introduction: Speech processing tasks, Speech science, language engineering applications.
2. Basic Tools: Analysis and spectral properties of the speech signal, linear prediction algorithms, statistical pattern recognition, dynamic programming, speech representation learning
3. Speech Coding: Human hearing properties, quantization theory, speech coding in the temporal and frequency domains.
4. Speech Synthesis: speech synthesis models, concatenative synthesis, hidden Markov models (HMMs) based speech synthesis, Neural speech synthesis.
5. Automatic Speech Recognition: Temporal pattern matching and Dynamic Time Warping (DTW) algorithms, speech recognition systems based on HMMs, Neural networks-based speech recognition.
6. Speaker recognition and speaker verification: Formalism, hypothesis testing, Text-dependent and Text-independent speaker verification, Gaussian mixture models-/HMM-based speaker verification, speaker embeddings based speaker verification, presentation attack detection (spoofing).
7. Paralinguistic speech processing: fundamentals and applications (e.g., emotion recognition, pathological speech detection, depression detection), hand-crafted feature based approaches, neural approaches.

Keywords

speech processing, speech coding, speech analysis/synthesis, automatic speech recognition, speaker identification, text-to-speech

Learning Prerequisites**Required courses**

Basis in linear algebra, signal processing (FFT), and statistics.

Important concepts to start the course

Basic knowledge in signal processing, linear algebra, statistics and stochastic processes. Basic knowledge

in machine learning/statistical pattern recognition is not a must but would be helpful in .

Learning Outcomes

By the end of the course, the student must be able to:

- Analyze speech signal properties
- Exploit those properties for speech coding, speech synthesis, speech recognition, speaker recognition and paralinguistic speech processing
- Formulate speech processing problems
- Choose appropriate methods for target speech processing tasks

Transversal skills

- Use a work methodology appropriate to the task.
- Access and evaluate appropriate sources of information.
- Use both general and domain specific IT resources and tools

Teaching methods

Lecture + lab exercises

Expected student activities

Attending courses and lab exercises. Read additional papers and continue lab exercises at home if necessary. Regularly answer list of questions for feedback.

Assessment methods

Written exam without notes

Resources

Bibliography

Fundamentals of Speech Recognition / Rabiner and Juang

Speech and Language Processing / Dan Jurafsky and Daniel Martin (2nd Edition)

Spoken language processing: A Guide to Theory, Algorithm and System Development / Xuedong Huang, Alex Acero and Hsiao-Wuen Hon

Speech and Audio Signal Processing: Processing and Perception of Speech and Music / Ben Gold, Nelson Morgan and Dan Ellis

Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing / Björn Schuller and Anton Batliner

Ressources en bibliothèque

- [Speech and Language Processing / Dan Jurafsky and Daniel Martin](#)
- [Fundamentals of Speech Recognition / Rabiner and Juang](#)
- [Spoken language processing: A Guide to Theory, Algorithm and System Development / Xuedong Huang, Alex Acero and Hsiao-Wuen Hon](#)
- [Speech and Audio Signal Processing: Processing and Perception of Speech and Music / Ben Gold, Nelson Morgan and Dan Ellis](#)
- [Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing / Björn Schuller and Anton Batliner](#)

Moodle Link

- <https://go.epfl.ch/EE-554>