



KTH Computer Science
and Communication

DT2118: Speech and Speaker Recognition

Course Analysis VT2013

1 Course data

Course name	Speech and Speaker Recognition
Course number	DT2118
Credits	7.5
When the course took place	Period 4, VT2013
Teachers (hours)	Giampiero Salvi (course responsible and lecturer, 18F, 2T, 4Ö, 8L, 4S) ¹
Number of registered students	13
Number of students attending	10
Number of students completing	8

2 Course objectives

The course objective is to give students insights in the signal processing and statistical methods used for speech and speaker recognition. After attending the course, the students should be able to:

1. discuss concepts related to speech signal processing and speech recognition
2. explain and implement simple parameter estimation methods for Gaussian models
3. explain and implement the main algorithms related to hidden Markov model training and decoding
4. use the software package KTH to build and evaluate a simple speech recogniser
5. carry out a small project related to speech or speaker recognition

3 How the course is designed to reach the objectives

The course gives a central space to the role of student activities as a means to learning. The 20 hours of lectures are meant to give the context and sufficient insights to the students in order to perform the different kinds of activities. Among these are: a set of computational exercises, a laboratory exercise and a final project.

The computational exercises require a mixture of theoretical and programming skills. The students are not only required to solve the exercises theoretically and numerically, but also to

¹F: lectures (föreläsningar), T: tutorial, Ö: exercises (övningar), L: lab, S: final seminar

hand in the code they have written for the solution. Matlab is the preferred language, but the students are left free to choose their favourite programming language. A central part of this activity has been the discussion with the teacher after the solution has been handed in. Given the relatively low number of students, the course responsible was able to sit with each student and check their solutions in details, searching for eventual bugs in their code.

The laboratory is based on a software package for speech recognition (HTK) and on recordings made by the students in a previous lab (from DT2112). Students work in groups of about 2. Students that did not participate in DT2112 are required to go through the DT2112 lab before they start the DT2118 lab. The aim of this exercise is to compare the effects of different feature extraction methods on the recognition accuracy in a digit recognition task. As an optional task, the students are asked to perform Vocal Tract Length Normalisation and to discuss the results they obtain.

The final activity is the final project. I tried to keep the subject of the projects as open as possible for two reasons. The first is the choosing the right project on the base of the information given by the lectures and one's own skills is in my opinion a learning activity *per se*. The other reason is that this kind of project work is best performed when the students are highly motivated, and being able to choose the task is a strong prerequisite to motivation. Of course, the students need continuous input from the teachers in order to realise if their ideas are feasible with the limited time given by the course. The students are required to work in groups of about 2 students (but also single student projects have been allowed this year). The need to perform the experiments related to the task they have chosen and write a report. Each student is also supposed to review the work of another group. In the final meeting in the course, each group is required to present the project, and the reviewers are required to ask related questions. Also the rest of the students are stimulated to participate in the discussion. In order to help the reviewers prepare for the discussion, each groups has been asked to hand in a draft of the report already one week before the final meeting.

4 Course pedagogical development I

This year I was responsible for all the activities in the course. Mats Blomberg and Kjell Elenius that were teaching assistants in previous years have retired. This gave me the opportunity to redesign many of the lectures making the course material more homogeneous. The lecture on advanced topics has also been updated to recent advances in the field of speech recognition.

The laboratory instructions have also been updated, and the scripts have been made compatible with the Ubuntu installations at CSC giving the students more freedom on where to perform the assignment.

The Course web pages have been moved to KTH Social, something that the students seem to have appreciated greatly.

5 Contact with the students during the course

5.1 Students in this year's committee

The student representatives for this year are Christos Kakouros and Emil Lundberg.

5.2 Results of the course meeting

There was no formal meeting during the course, but there was always close discussion with the students, also simplified by the low number of students.

6 Contact with the other teachers during the course

This point is not relevant being I the only teacher.

7 Questionnaire, the student's point of view

7.1 Period in which the questionnaire was active

After the last meeting of the course, until the final meeting with the course responsible (about two weeks)

7.2 Questions that were added to the standard

See the questionnaire

7.3 Statistics of answers

50% of the active students answered the questionnaire.

7.4 Changes compared to the last implementation

No modifications were made this year.

7.5 General impression

The general impression on the course is positive. Students appreciate especially the computational exercises that give insights in the details of the algorithms, and the fact that the teacher was always available for discussion (this is of course dependent on the number of students that enrolled this year).

Among the written comments there is some constructive criticism, but always combined with positive forced choice answers.

7.6 Relevant web links

8 Interpretation of the questionnaire by the course responsible

8.1 Positive views

- Most students had a clear idea of the course objectives from the start.
- All students this year are happy about the balance between standard lectures and practical activities.
- Students are very happy about the computational exercises.
- Students liked in general the availability and engagement of the teachers
- Students are happy about the organisation of the course and the new KTH Social platform.

8.2 Negative views

- Many students did not use the course book, or had some criticism about it.
- One student complained he or she had not a clear idea of how much study time was requested for the final project.
- Many students expressed the wish for exercises in class.
- one student found the lectures and activities a bit disconnected.

8.3 Was the course relevant with respect to the objectives

Yes

8.4 View on prerequisites

The prerequisite of DT2112 should be removed as it is not necessary for this course. Some students were not aware of it and many did not attend DT2112 and completed DT2118 successfully.

8.5 View on forms of teaching

It seems that the lecturers were considered positively from a pedagogical perspective

8.6 View on literature and course material

There is always some student criticising the book. Not sure how many feel that way. The rest of the material is appreciated, especially the fact that I was giving handouts before the lectures.

8.7 Opinion on exam

Most students are positive about the project presentations

8.8 Especially interesting comments

- “Would prefer that there would be no course book.”
- “Very active KTH Social page — good!”
- “Thanks for the course! Really content with the time invested in it.”

9 Opinions from the other teachers after the end of the course

9.1 What worked well

9.2 What worked less well

9.3 Suggestions for improvements

10 Results of the course commission meeting after the exam

10.1 Summary from the students

10.2 Suggestions for changes

10.3 Link to the meeting protocol

11 Summary of the course responsible report

11.1 General impression

The general impression of the course is positive.

11.2 Positive points of view

There seems to be a good balance between lectures and practical assignments. The availability of the teachers seems to be highly appreciated by the students. The organisation of the different steps in the course seems to have worked well. The students were very motivated, they attended the lectures regularly (although it was not mandatory) and participated actively in the discussions. They respected all the deadlines, sometimes anticipating them.

11.3 Negative points of view

The lectures are a bit too much concentrated at the beginning of the course. Although this was done on purpose, it seems to make it harder for the students to elaborate the concepts that are discussed in the theoretical part of the course.

11.4 Opinion on prerequisites

The prerequisite are satisfactory. Some details will be adjusted.

11.5 Opinion on forms of teaching

The forms of teaching seem to be satisfactory (see also previous points)

11.6 Opinion on literature/course material

The literature seems to be satisfactory.

11.7 Opinion on examination

The examination seems to have worked fine.

12 Course pedagogical development II

12.1 How the changes to this course work

The changes to the previous years seem to have worked well. Especially the decision to check the exercise solutions individually with the students.

12.2 Changes that should be done for the next time

- some of the lectures will be moved towards the end of the course.
- some ready made material for the project will be made available in order to simplify the work
- we will consider the possibility of designing in class exercises
- the lab will be modified to reduce the running time. It is also desirable to add some intermediate steps in the lab so to get the students more involved in the details of its implementation.

DT2118 Speech and Speaker Recognition

Resultat av kursutvärdering

25 questions to be answered and a few comments to be given.

1. Do you think the course is easy or difficult?

1. 0% (0 st) Very easy.
 2. 20% (1 st) Easy.
 3. 20% (1 st) Average.
 4. 60% (3 st) Rather difficult.
 5. 0% (0 st) Very difficult.
-

2. Did you get a clear idea of the course objective at the course start?

1. 60% (3 st) Yes.
 2. 40% (2 st) Hesitant.
 3. 0% (0 st) No.
-

3. Do you think the course is interesting?

1. 40% (2 st) Yes, very.
 2. 60% (3 st) Yes.
 3. 0% (0 st) Neutral.
 4. 0% (0 st) Not very.
 5. 0% (0 st) No.
-

4. How did you get to know about this course?

1. 0% (0 st) Friend.
2. 40% (2 st) KTH web pages.
3. 40% (2 st) From DT2112.
4. 20% (1 st) Other (please specify).

Comments:

I think I had heard about it before, but I basically took it into serious consideration through lab 1 of DT2112 that mentioned that the results would be used for a follow up course.

Mandatory course.

5. The requirement on previous courses is the Speech Technology course DT2112 or equivalent and experience with unix. Do you regard your level as sufficient at the time of the course start?
1. 40% (2 st) Yes.
 2. 60% (3 st) Hesitant.
 3. 0% (0 st) No.

Comments:

DT2112 gave very profound background. Unix experience to the level we needed it for the lab was not too high or overwhelming.

I actually didn't know about the Speech Technology requirement...

Didn't take DT2112 or equivalent, initial theory was hard to understand.

6. What is your opinion on the course book "Spoken Language Processing"?
1. 0% (0 st) Very good.
 2. 40% (2 st) Good.
 3. 20% (1 st) Acceptable.
 4. 0% (0 st) Not so good.
 5. 0% (0 st) Bad.
 6. 40% (2 st) Did not use it.

Comments:

I used Martin - Jurafsky book also. The course book is ok. Sometimes too thin in the descriptions, sometimes too detailed and may contain some errors.

Would prefer that there would be no course book.

7. What is your opinion on the extra material (papers, handouts, etc)?
1. 20% (1 st) Very good.
 2. 80% (4 st) Good.
 3. 0% (0 st) Acceptable.
 4. 0% (0 st) Not so good.
 5. 0% (0 st) Bad.
 6. 0% (0 st) Did not use it.

Comments:

Pretty informative on the areas covered. They gave sufficient hints for the subjects. Studying the book

was mandatory though.

Handouts were generally good, liked the idea of giving them out to students at the beginning of each lecture.

8. Was there a good balance between lectures and practical activities?

1. 0% (0 st) Too many lectures.
2. 100% (5 st) Good balance.
3. 0% (0 st) Too many practical activities.

Comments:

Lab was really nice. The exercises too. The lectures were just as many as it should be. Not too many, not too little.

I feel that lectures and activities were a bit disconnected.

9. What do you think of the lectures in a pedagogical way? (Are the concepts well described? Do the teachers speak and write clearly)

1. 20% (1 st) Very good.
2. 80% (4 st) Good.
3. 0% (0 st) Acceptable.
4. 0% (0 st) Not so good.
5. 0% (0 st) Bad.

Comments:

Sometimes, the teacher would speak too much without interaction from the audience. This might be a problem of the audience.

Especially the first lecture with the live demo of vowels' anatomy (how the spectra look) etc.

Missed most lectures, but judging from the ones I attended the teacher is very responsive and eager to explain more if needed.

10. Do you think the theoretical aspects should be expanded more in details?

1. 0% (0 st) Fully agree.
2. 20% (1 st) Agree.
3. 80% (4 st) Do not have an opinion.
4. 0% (0 st) Disagree.
5. 0% (0 st) Fully disagree.

Comments:

Some aspects are a bit hard to get with a simple lecture. Maybe some exercises in class after the corresponding lecture as a session with teacher, would help get them easier.

Maybe as bonus material, for the benefit of students who are more interested in the topic.

11. How was your work load distributed during the course?

1. 0% (0 st) Mainly at the beginning.
2. 80% (4 st) Evenly during the course.
3. 20% (1 st) Mainly at the end.

Comments:

Not that much in the first two weeks. Then the lab came, the exercises and the project. So apart from the first 2 weeks, the rest was even.

I feel that students were expected to be very responsible with their time management, something that should maybe be made clearer.

12. The assessment in this course is based on a Fail/Pass grading scale. Do you think we should assign more specific grades? For example on a 7 level scale?

1. 20% (1 st) Yes.
2. 20% (1 st) Hesitant.
3. 60% (3 st) No.

Comments:

I do not know. I took the course not only because of the subject, but also because of the fact that it has practical work and not just a written exam in the end of the course. If this remains maybe a 7 scale can be introduced. I liked it this way though.

Fail/Pass seems to work fine.

13. What is your opinion on the computational exercises?

1. 40% (2 st) Very good.
2. 60% (3 st) Good.
3. 0% (0 st) Acceptable.
4. 0% (0 st) Not so good.
5. 0% (0 st) Bad.
6. 0% (0 st) No opinion.

Comments:

Especially, the last ones on HMMs. They also pointed computational efficiency aspects (like the fact that we use logarithms to avoid underflow)

Good at ensuring students have at least some basic experience with the basic ideas used.

14. What is your opinion on the practical exercise (lab)?

1. 40% (2 st) Very good.
2. 60% (3 st) Good.
3. 0% (0 st) Acceptable.
4. 0% (0 st) Not so good.
5. 0% (0 st) Bad.
6. 0% (0 st) No opinion.

Comment:

Maybe the student should be asked to take a look at the scripts and describe how they work. We get a lot for free in these labs and only when I studied the scripts did I fully get the concepts and how it performs the recognition tasks.

I would like to see a description of the expected workload for that, such as the expected amount of time it will take, or the expected amount of studying the manual one has to do.

15. How long time did you spend preparing the practical exercise (lab)?

1. 0% (0 st) Less than 6 hours.
2. 40% (2 st) 6-12 hours.
3. 60% (3 st) 12-24 hours.
4. 0% (0 st) More than 24 hours.

Comment:

Running and getting results took us like 4-5 hours for lab 1 and another 4-5 hours for lab 2. We took 3-4 hours writing the reports and some 2 hours to go through the scripts.

A bit hard to measure, since it was a group assignment.

16. How long time did you spend on writing the term paper and preparing the presentation:

1. 0% (0 st) Less than one day.
2. 0% (0 st) 2-3 days.
3. 60% (3 st) 4-6 days.

4. 40% (2 st) 7 days or more.

Comments:

Not working fully all day long with the project. We seriously started it 2 weeks before the presentation. During the last week we devoted more time in it.

Term paper was a continuous process, hard to measure.

17. How many other courses did you follow in parallel to this one (period 4)?

1. 0% (0 st) None.
 2. 40% (2 st) One.
 3. 60% (3 st) Two.
 4. 0% (0 st) Three.
 5. 0% (0 st) Four or more.
-

18. How large proportion of your studying time in period 4 did you spend on this course?

1. 0% (0 st) Less than 15%.
 2. 0% (0 st) 15-30%.
 3. 80% (4 st) 30-50%.
 4. 20% (1 st) 50-70%.
 5. 0% (0 st) More than 70%.
-

19. The course is 7.5 hp. How do you regard that in comparison with other courses?

1. 100% (5 st) 7.5 hp is appropriate.
 2. 0% (0 st) Should be more than 7.5 hp.
 3. 0% (0 st) Should be less than 7.5 hp.
-

20. What is your opinion on the administration of the course and exchange of information between teachers and students?

1. 60% (3 st) Very good.
2. 40% (2 st) Good.
3. 0% (0 st) Acceptable.
4. 0% (0 st) Not so good.
5. 0% (0 st) Bad.

Comments:

The teacher organized private discussions, in the lectures always asked for possible questions we might have and always stressed out that we can contact him for questions.

Very active KTH Social page - good!

Didn't have much communication, was otherwise good.

21. Were the requirements at every stage of the course always clear to you?

1. 80% (4 st) Yes.
2. 0% (0 st) Hesitant.
3. 20% (1 st) No.

Comments:

Maybe concerning the project, it was not that clear how much time we needed to devote in it and how much of the course it occupies.

Had to backtrack into the theory during the assignments here and there.

22. Do you feel that you have been discriminated in this course due to gender, sexuality, ethnicity or disability?

1. 0% (0 st) Yes.
2. 0% (0 st) Hesitant.
3. 100% (5 st) No.

If yes, in which way?

23. How do you regard the course from a gender perspective (e.g. with respect to course book, teachers, etc.)?

I do not quite get this question... :(

I did not feel that anyone was discriminated or anything.

Mostly men seem to be interested in all courses at TMH, no other gender thoughts than that.

All was fine with regard to that.

24. Suggestion for course improvements:

Some exercises in class could be added in these topics that can support exercises. Students could be asked to solve them and then they could be presented and discussed with each other and the teacher (like a seminar).

Avoid the need for a book! Also, a short summary of the theory needed before starting each assignment would be nice.

25. Further comments:

Thanks for the course! Really content with the time invested in it.

I'm a big fan of having a project instead of a traditional exam, since exams don't really feel like they test relevant skills in this kind of course. To actually apply what you've learned in a concrete project is a lot better in my opinion.

All in all a good course, the final presentation time was fun as well.

giampi@kth.se

Denna sammanställning har genererats med [ACE](#).

Student Meeting Minutes (2013-06-05)

Participants: Giampiero Salvi, Christos Kakouros and Emil Lundberg

Penna: Christos Kakouros

General remarks

- Course seemed surprisingly difficult for 3 out of 5 students that replied to the questionnaire. Reasons why the course was difficult are not very clear: all the students completed the assignments successfully.
- The course was radically changed from previous years when it was addressing PhD students. Many lectures were redesigned.
- It is not mandatory to have taken DT2112 Speech Technology to attend the course or the other way round.
- The students can manage the course without following DT2112, thus the latter is not a prerequisite.
- One of the difficulties of many students was their lack in background in signal processing. This is also associated with general speech concepts that might not be well understood. Some students might learn how to do something but not really why they do it. Devoting more time in it during the first lectures would probably give a better understanding of the speech related concepts. Adding some exercises on these topics, apart from the computational ones could also be suggested.
- Related to the above, the course deepens a lot in practical skills which is favored by the students but maybe more theory could be added.
- The language part of the course is not very broad (language models, grammars, etc.). Some exercises could be added on that topic but then maybe should be considered optional because the extent of the exercises would grow a lot.
- Extra exercises sessions could be added in the lectures. These could cover topics like N-grams or other topics not covered by the current computational exercises. They should also not be too related to large data since they cannot really be done in the context of a classroom session but using computers instead.
- The amount of time that the student is expected to devote per part of the course (Lab, Exercises, Project) could be clarified a bit during the first lectures.

Book

- Some complaint was expressed for the book. Another book could be used or there could be an option for the student between the current and some other books.

Labs

- Now there is enough speech data to change the second lab. Leave-one-out validation sets might not be required anymore.
- Another point that may cause some students to learn how to do things but not why, could be the fact that they are given too much for free in the labs.
 - Matlab could be an alternative where students could be given some code with the key part missing and then asked to fill it.

- However, HTK is an established tool for speech recognition and using it for the lab gives a nice background for students who choose theses in TMH (Department of Speech, Music and Hearing).
- HTK is open-source but it is difficult to let the students fill in code parts (like the proposed solution with Matlab above) since they could be easily overwhelmed by the lengthy code of the existent HTK functions. What could work instead for that case would be to add comments to the existing scripts of the lab and ask the students to fill in table 1 (page 7) of lab 1 description for example.