Spraakherkenning en synthese 2015 (onder voorbehoud)

David Weenink

January 6, 2015

Administrativa

Classes and practicum: Tuesday, Wednesday, Thursday, Friday

Tentamen: Friday January 30, 9-12

Hertentamen:


1 Introduction

This is a preliminary overview of the contents of this course and the contents if this document might change during the development of the course.

We will use the books Daniel Jurafsky and James H. Martin, *Speech and Language Processing* 2nd (Prentice Hall, 2009) and Weenink (2015), *Speech signal processing by Praat* (available at [http://www.fon.hum.uva.nl/david/sspbook/sspbook.pdf](http://www.fon.hum.uva.nl/david/sspbook/sspbook.pdf)) and a number of articles as theoretical background, while the computer program Praat (available at [http://www.praat.org](http://www.praat.org)) will be used for the practical clarification of many of the introduced concepts.

1.1 Course content

This course will contain five parts:

1. Introduction to speech sounds and their analysis. Various aspects of speech sounds will be introduced like the impressionistic classification of speech sounds, the source-filter model of speech production and various analyses of the speech sound signal.

2. Speech synthesis. Here different speech synthesis models will be introduced like formant, diphone and unit synthesis. Also the things needed to go from text to speech will be treated.

3. Speech recognition. Dynamic Time Warping and Hidden Markov models will be introduced as the building blocks used in different kinds of speech recognition and synthesis systems.


5. The practicum where some of the techniques introduced during the classes will be used. We advise you to carry your own headphones because you will have to listen to sounds a lot.

1.2 Examination

The course will be terminated by an examination on Friday, January 30 from 9-12. This examination will determine your final grade. You can only attend the exam after all of your assignments have been fulfilled. If all your assignments have been graded as “voldoende” you will earn an extra point for the exam.

1.3 The practicum

During the practicum you make your assignments and you will receive feedback on them whether they are “voldoende” or “onvoldoende”. A “voldoende” will always be graded as 6 an “onvoldoende” always as 5. Some assignments are easy, some involve more work. I will give groupwise feedback on the assignments. If after the explanation you still have questions or want feedback or further clarification on the way you solved the assignment you can ask me during these hours. I will always be present during practicum hours.

As many assignments involve listening to sounds you should bring your own earphones (and your own laptop too, if you like).

Please send the assignments to BB before the next practicum session.
2 Week 1, Jan 6-9: The speech signal

2.1 The speech signal, classification of speech sounds

- Speech production: the speech organs; speech perception: the ear
- Speech sounds: vowels and consonants, the oscillogram
- Classification of sounds: place, manner and voiced/unvoiced
- Introduction to Praat


2.2 Sampling and quantization, Praat scripting

- Acoustics, Sampling and quantization, Nyquist theorem
- Signal reconstruction from the samples
- Praat scripting

- Literature: Weenink [2015] Ch. 3,4

2.3 Spectrum and spectrogram

- Fourier analysis and spectra
- Difference between oscillogram, spectrum, spectrogram

- Literature: Jurafsky and Martin [2009] ch. 7.4, Weenink [2015] Ch. 7,8

2.4 Feedback and loose ends

Feedback on assignments and things that have been lost.

3 Week 2, Jan 13-16: Speech analysis & synthesis

3.1 Pitch

- What is pitch?
- How to measure
- Autocorrelation function
- Intonation
- Annotation

- Literature: Boersma [1993], Weenink [2015] Ch. 5

3.2 Class 4 Digital Filtering & Convolution

- Digital filter, impulse response, convolution
- Formant filter, antiformant filter

- Weenink [2015] Ch. 11

3.3 LPC & Formant frequency measurements

- What are formants?
- Linear prediction coefficients
- Z-transform

- Literature: Markel and Gray [1976], Makhoul [1975], Weenink [2015] Ch. 16
3.4 Feedback and loose ends
Feedback on assignments and things that have been lost.

4 Week 3, Jan 20-23: Speech recognition

4.1 Synthesis of vowels
- Formant synthesis of vowels
- VowelEditor to show synthesis of vowels

4.2 Text To Speech

4.3 MFCC & DTW & Speak Good Chinese
- Spectral representations: Mel frequency cepstral coefficients
- Dynamic time warp
- Literature: Sakoe and Chiba [1978] Example of elementary techniques
- Literature: Weenink et al. [2007]

4.4 Feedback & Corpora
- Literature: Jurafsky and Martin [2009] ch. 7.5

5 Week 4, Jan 27-30: Speech recognition

5.1 Markov models, Gaussian mixture models
- Literature: Rabiner [1989], Jurafsky and Martin [2009] ch. 6,9

5.2 Application of Markov models
- Literature: Jurafsky and Martin [2009] ch. 6,9

5.3 Separation of voices
- Blind source separation, Independent component analysis
- Cross-correlation
- Literature: Hyvärinen and Oja [2000], Ziehe et al. [2004]

5.4 9-12 Tentamenstof
- Classes (sheets)
- Assignments
- Jurafski & Martin, chapters 6, 7.1–7.5, 8, 9
- Papers Rabiner [1989], Sakoe and Chiba [1978], Weenink et al. [2007]
References


