

# Spraakherkenning en synthese 2014 (onder voorbehoud)

David Weenink

January 3, 2014

## Administrativa

**Classes and practicum:** Monday, Tuesday, Wednesday, Friday

**Tentamen:** Friday January 31, 13-16

**Hertentamen:**

**SHS 2013 website:** [http://www.fon.hum.uva.nl/david/ba\\_shs/2014](http://www.fon.hum.uva.nl/david/ba_shs/2014)

## 1 Introduction

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

We will use the books Jurafsky and Martin [2009], Weenink [2014] and a number of articles as theoretical background, while the computer program Praat [Boersma and Weenink, 2013] 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.
4. Miscellaneous. Speech Corpora and their use speech technology.
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 31 from 13-16. 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 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 and “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 6-10: 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
- Literature: Jurafsky and Martin [2009, ch 7.1-7.3], Weenink [2014, Ch. 1, 2]

### **2.2 Sampling and quantization, Praat scripting**

- Acoustics, Sampling and quantization, Nyquist theorem
- Signal reconstruction from the samples
- Praat scripting
- Literature: Weenink [2014, Ch. 3,4]

### **2.3 Spectrum and spectrogram**

- Fourieranalysis and spectra
- Difference between oscillogram, spectrum, spectrogram
- Literature: Jurafsky and Martin [2009, ch. 7.4], Weenink [2014, Ch. 7,8]

### **2.4 Feedback and loose ends**

Feedback on assignments and things that have been lost.

## **3 Week 2, Jan 13-17: Speech analysis & synthesis**

### **3.1 Pitch**

- What is pitch?
- How to measure
- Autocorrelation function
- Intonatie
- Annotation
- Literature: Boersma [1993], Weenink [2014, Ch. 5]

### **3.2 Class 4 Digital Filtering & Convolution**

- Digital filter, impulse response, convolution
- formant filter, antiformant filter
- Weenink [2014, Ch. 11]

### **3.3 LPC & Formant frequency measurements**

- What are formants?
- Linear prediction coefficients
- Z-transform
- Lit: Markel and Gray [1976], Makhoul [1975], Weenink [2014, ch. 16]

### **3.4 Feedback and loose ends**

Feedback on assignments and things that have been lost.

## **4 Week 3, Jan 20-24: Speech recognition**

### **4.1 Synthesis of vowels**

- Formant synthesis of vowels
- VowelEditor
- Literature: Klatt [1980], Klatt and Klatt [1990], Weenink [2014, ch. 12]

### **4.2 Text To Speech**

- Literature: Jurafsky and Martin [2009, ch. 8],

### **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-Jan 31: 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
- Lit:Hyvärinen and Oja [2000], Ziehe et al. [2004]

### **5.4 13-16 Tentamenstof**

- Classes (sheets)
- Assignments
- Weenink [2014, chapters 1– 13, 16, Appendix A,B]
- Jurafski & Martin, chapters 6, 7.1–7.5, 8, 9
- Papers Rabiner [1989], Sakoe and Chiba [1978], Weenink et al. [2007]

## References

- Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc. Institute of Phonetic Sciences University of Amsterdam*, 17:97–110, 1993.
- Paul Boersma and David J. M. Weenink. Praat: doing phonetics by computer [computer program], 2013. URL <http://www.praat.org/>.
- A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, 2000.
- Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Prentice Hall, 2nd edition, 2009.
- Dennis H. Klatt. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.*, 67:971–995, 1980.
- Dennis H. Klatt and Laura C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, 87:820–857, 1990.
- J. Makhoul. Linear prediction: A tutorial review. *Proc. IEEE*, 63:561–580, 1975.
- J. D. Markel and A. H. Gray, Jr. *Linear prediction of speech*. Springer Verlag, Berlin, 1976.
- L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.
- Hiroaki Sakoe and Seibi Chiba. Dynamic programming optimization for spoken word recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 26:43–49, 1978.
- David Weenink, G. Chen, Z. Chen, S. de Konink, D. Vierkant, E. van Hagen, and R.J.J.H. van Son. Learning tone distinctions for mandarin chinese. *Proceedings Interspeech*, pages 950–953, 2007.
- David J. M. Weenink. *Speech signal processing by Praat*. To be published, 2014. URL <http://www.fon.hum.uva.nl/david/sspbook/sspbook.pdf>.
- Andreas Ziehe, Pavel Laskov, Guido Nolte, and Klaus-Robert Müller. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation". *Journal of Machine Learning Research*, 5:777–800, 2004.