

Practicum Taal&Sprak

1e deel: (spraak)signalen

<http://www.ai.rug.nl/nl/vakinformatie/PTenS/>

Tjeerd Andringa
tjeerd@ai.rug.nl

Maria Niessen
maria@ai.rug.nl



university of
 groningen

faculty of mathematics
and natural sciences

artificial
intelligence

auditory
cognition

Even voorstellen

Maria Niessen

- Oud KI-er
- Promovendus (geluidsbron herkenning)
- iCat laten luisteren
- Taken
 - Coördinatie cursus
 - Begeleiding practica
 - Nakijken opgaven
 - Probleempjes bij practica



Tjeerd Andringa

- Ooit vaste stof fysica
- Ooit kernstaf TCW
- OWD-KI
- Onderzoek
 - Spraak/geluid herkenning
 - Auditory Scene Analysis
 - Auditory Cognition
 - Link fysica en geluid
- Bedrijf: Sound Intelligence
 - onderzoeksprojecten



artificial
intelligence

auditory
cognition

2

Auditory Cognition Group

Group leader:

- Tjeerd Andringa
 - Director of education AI

Postdoc:

- Ronald van Elburg
 - Directional hearing
 - Computational Neuroscience

Advisor:

- Peter van Hengel (Sound Intelligence)

Origins of ACG:

- Sensory biophysics
 - Duijhuys (>1980), basic cochlea model
- Artificial Intelligence (>1991)
 - Best AI-programme in NL
- Sound Intelligence
 - Important basic research

PhD-projects:

- Dirkjan Krijnders: Multi-media aggression detection (Cassandra)
- Renante Vieland: Linking physics to sounds
- Hedde van de Vooren: Condition Monitoring
- Maria Niessen: Everyday sound recognition
- Bea Valkenier: Key word spotting

Master Projects:

- Maarten van Grootel: environmental sounds
- Carina Pals: Auditory object perception
- Henrik Kjeldsen: Holographic neural networks



university of
 groningen

faculty of mathematics
and natural sciences

artificial
intelligence

auditory
cognition

3

Eerste college

- Onderwerpen en doelen collegereeks
- Vorm
- Relevantie
- Praktische zaken
 - Indeling in tweetallen (pauze)
 - Verdeling 2e hoorcollege
- Voorbereiding eerste practicum (vanmiddag)



university of
 groningen

faculty of mathematics
and natural sciences

artificial
intelligence

auditory
cognition

4

Vorm

- Hoorcolleges
 - In dienst van practica
 - Theoretische achtergronden
 - Voorbereidingen op het practicum
 - Rapportage over practicumopdrachten.
 - Zowel docenten als studenten
- Computerpractica
 - Toetsen (toegangscontrole)
 - kookboek practica met vragen
 - kleine projecten
- Deeltentamens (3x)
- Hertentamen (1x, mondeling)



university of
 groningen

faculty of mathematics
and natural sciences

artificial
intelligence

auditory
cognition

5

Eindcijfer

Practica verplicht
anders geen eindcijfer!

Tentamencijfer:

- 50%
 - Toetsen
 - Practica
 - Presentaties
- 50%
 - Deeltentamens (15% - 20% - 15%)

Hertentamen → Mondeling!



university of
 groningen

faculty of mathematics
and natural sciences

artificial
intelligence

auditory
cognition

6

Tweetallen

Individueel

- Toetsen
- deeltentamens

Gezamenlijk

- Practica
- Presentaties

Groepjes tijdens pauze melden bij student-assistent

Overzicht collegereeks

1 Algemene signaalanalyse

- Fouriertransformatie, FFT
- Spectrale analyse

2 Spraaksignaalanalyse

- Geluidsopnamen, sampling
- Spraakklanken, Formanten, toonhoogte

3 Automatische Spraakherkenning (ASR)

- Trainen en testen ASR systemen
- Mogelijkheden en beperkingen, ruis

4 Continuity Preserving Signal Processing

- Signaal-ruis verhouding
- Spontane spraak
- Herkenning in ruis

Onderwerpen en doelen

Onderwerpen

- (Spraak)signaalanalyse
- Spraaksignaal eigenschappen
- Automatische spraakherkenning
- Cognitiewetenschappelijke bijdragen
 - Continuity Preserving Signal Processing (CPSP)
 - Auditory Scene Analysis (ASA)
 - Spontane/emotionale spraak
- Toepassingen
 - Selectie van spraak uit ruis

Doelen:

- Overzicht spraaktechnologie
- Gevoel voor:
 - Complexiteit
 - Waarom spraak is zoals het is
 - Rol van ruis
 - Redundantie in signaal
 - Beperkingen ASR-systemen
 - Waar liggen mogelijkheden
 - En waar niet

Wetenschappelijke context

Cognitive Systems: a new research goal

FP7, ICT-challenge 2:

The increasing complexity of our society and economy places greater emphasis on artificial systems such as robots, smart devices and machines which can deal autonomously with our needs and with the peculiarities of the environments we inhabit and construct. This challenge is to extend systems engineering methods to deal with open-ended and frequently changing real-world environments.

EU program manager:

*"Stop making toy-systems.
We know you can do that.
Impress us outside the lab!"*

Engineering vs natural system



Operating environment of modern ASR systems

- simple (one source)
- known
- controlled
- constant
- limited

Operating environment human sound recognition

- complex (many sources)
- unknown
- uncontrolled
- variable
- unlimited

Large number of task and domain specific systems

One system for all tasks and all environments

Relevantie

- Automatische spraakherkenning
 - Dicteesystemen
 - Dialogsystemen
 - Keyword spotting
- Spreker herkenning
 - Identificatie
 - Verificatie
- Scheiden en herkennen van geluiden
 - Verschillen tussen spraak, muziek en de rest?
 - Wat is (potentieel) interessant en wat (dus) niet
 - Wat gebeurt er in de omgeving?

Voorbereiding practicum 2

- Studenten hoorcollege

 - Max 15 min
 - Kern: geen details
 - Van te voren doorspreken met Tjeerd (bij practicum)
 1. Gebruik opname programmatuur.
 - dynamisch bereik
 - sample frequenties
 - audiocoderingstechnieken

2. MP3
 - psychoacoustiek
 3. Toonhoogte bepaling
 - harmonisch nummer
 4. LPC (linear predictive coding)
 - Telefooncodering
 5. MFCC
 - ASR-input

1^e practicum (twee weken)

- Goed voorbereiden!
 - (Eerste 10 minuten toets)
 - Tweetalen
 - Computerzaal: Bernoulliborg 283
- Onderwerpen

 - Basisvectoren
 - Discrete Fourier-transformatie
 - Complexe Fourier-transformatie: FFT
 - Efficiency van de FFT
 - Impulsrespons
 - Windows (vensters)
 - Logsweep
 - Spectrogram
 - Zichtbaar maken van signaal componenten

Signaalanalyse: Bekend terrein!

- Werken met rijtjes getallen (vectoren)
- Vector meetkunde (inproducten, hoeken)
 - Statistiek (correlaties)
 - Neurale netwerken (gewogen input: $f(\mathbf{x}) \cdot \mathbf{w}_{ij}$)

- Vergelijkbaar
- onderling loodrecht: $\mathbf{x} \cdot \mathbf{y} = 0$
 - hoek tussen vectoren: $\cos(\angle(\mathbf{x}, \mathbf{y})) = \mathbf{x} \cdot \mathbf{y} / (|\mathbf{x}| |\mathbf{y}|)$
 - afhankelijkheid, $\mathbf{x}$ en $\mathbf{y}$ in zelfde richting: $\mathbf{x} = a \mathbf{y}$
 - gecorreleerd: $\mathbf{x}$ zegt iets over $\mathbf{y}$ d.w.z. $\text{cor}(\mathbf{x}, \mathbf{y}) \approx 1$

Transformeren = afbeelden op basisvectoren

- Basis $\mathbf{p}$
- $\{\mathbf{p}_1, \mathbf{p}_2\} = \{(1,0), (0,1)\}$
- Basis $\mathbf{b}$
- $\{\mathbf{b}_1, \mathbf{b}_2\}$

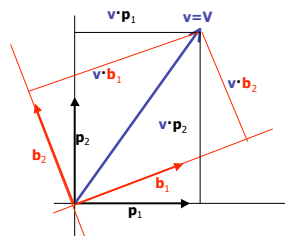
Vector $\mathbf{v}$ op $\mathbf{p}$ en $\mathbf{V}$ op $\mathbf{b}$

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{p}_1) \mathbf{p}_1 + (\mathbf{v} \cdot \mathbf{p}_2) \mathbf{p}_2$$

$$\mathbf{V} = (\mathbf{V} \cdot \mathbf{b}_1) \mathbf{b}_1 + (\mathbf{V} \cdot \mathbf{b}_2) \mathbf{b}_2$$

Basis $\mathbf{b}$

$$\mathbf{b}_1 = (\mathbf{b}_1 \cdot \mathbf{p}_1) \mathbf{p}_1 + (\mathbf{b}_1 \cdot \mathbf{p}_2) \mathbf{p}_2$$

$$\mathbf{b}_2 = (\mathbf{b}_2 \cdot \mathbf{p}_1) \mathbf{p}_1 + (\mathbf{b}_2 \cdot \mathbf{p}_2) \mathbf{p}_2$$


Fourier basisvectoren (dim=4)

- Tijddomein
- Basis: pulsen
 - Geen frequentie informatie

- Frequentiedomein
- Basis: golven
 - Geen tijdsinformatie

$\mathbf{b}_1 = [1 \ 0 \ 0 \ 0]$
 $\mathbf{b}_2 = [0 \ 1 \ 0 \ 0]$
 $\mathbf{b}_3 = [0 \ 0 \ 1 \ 0]$
 $\mathbf{b}_4 = [0 \ 0 \ 0 \ 1]$

$\mathbf{B}_1 = \cos(\omega_0) = [1 \ 1 \ 1 \ 1]$
 $\mathbf{B}_2 = \cos(\omega_1) = [1 \ 0 \ -1 \ 0]$
 $\mathbf{B}_3 = \sin(\omega_1) = [0 \ 1 \ 0 \ -1]$
 $\mathbf{B}_4 = \cos(\omega_2) = [1 \ -1 \ 1 \ -1]$
 $\sin(\omega_0) = \sin(\omega_2) = [0 \ 0 \ 0 \ 0]$

Basis vectoren: onderling loodrecht

Tijddomein

$$b_2 \cdot b_3 = \text{sum}([0 \ 1 \ 0 \ 0]) \cdot [0 \ 0 \ 1 \ 0] = \text{sum}([0 \ 0 \ 0 \ 0]) = 0$$

$$b_2 \cdot b_2 = \text{sum}([0 \ 1 \ 0 \ 0]) \cdot [0 \ 1 \ 0 \ 0] = \text{sum}([0 \ 1 \ 0 \ 0]) = 1$$

Norm = 1

Frequentiedomein

$$\cos(\omega_1) \cdot \sin(\omega_1) = \text{sum}([1 \ 0 \ -1 \ 0]) \cdot [0 \ 1 \ 0 \ -1] = 0$$

$$\cos(\omega_1) \cdot \cos(\omega_1) = \text{sum}([1 \ 0 \ -1 \ 0]) \cdot [1 \ 0 \ -1 \ 0] = 2$$

$$= \sqrt{2} \cdot \sqrt{2}$$

$$\sin(\omega_1) \cdot \sin(\omega_1) = \text{sum}([0 \ 1 \ 0 \ -1]) \cdot [0 \ 1 \ 0 \ -1] = 2$$

$$= \sqrt{2} \cdot \sqrt{2}$$

$$\cos(\omega_0) \cdot \cos(\omega_0) = \text{sum}([1 \ 1 \ 1 \ -1]) \cdot [1 \ 1 \ 1 \ -1] = 4$$

$$= 2 \cdot 2$$

$$(B) = \begin{Bmatrix} \cos(\omega_0) & \cos(\omega_1) & \sin(\omega_1) & \cos(\omega_2) \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{2} \end{Bmatrix}$$



Representatie van s=[2 -1 0 -1]

Tijddomein:

$$s_1 = s \cdot b_1 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [1 \ 0 \ 0 \ 0] = 2$$

$$s_2 = s \cdot b_2 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [0 \ 1 \ 0 \ 0] = -1$$

$$s_3 = s \cdot b_3 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [0 \ 0 \ 1 \ 0] = 0$$

$$s_4 = s \cdot b_4 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [0 \ 0 \ 0 \ 1] = -1$$

Frequentiedomein:

$$S_1 = s \cdot b_1 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [1 \ 1 \ 1 \ 1]/4 = 0$$

$$S_2 = s \cdot b_2 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [1 \ 0 \ -1 \ 0]/2 = 1$$

$$S_3 = s \cdot b_3 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [0 \ 1 \ 0 \ -1]/2 = 0$$

$$S_4 = s \cdot b_4 = \text{sum}([2 \ -1 \ 0 \ -1]) \cdot [1 \ -1 \ 1 \ -1]/4 = 1$$

FT van s in tijddomein is S=[0 1 0 1] in frequentiedomein

Want: $s = S_2 \cdot b_2 + S_4 \cdot b_4 = [1 \ 0 \ -1 \ 0] + [1 \ -1 \ 1 \ -1]$
 $= [2 \ -1 \ 0 \ -1]$



FT als matrix vermenigvuldiging

Dim=4

Inproduct = product: rijvector x kolomvector

$$S = [s \cdot b_1 \ s \cdot b_2 \ s \cdot b_3 \ s \cdot b_4]$$

$$B = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -1 \end{bmatrix} \quad B' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

$$\text{Heen: } s = s \cdot B$$

$$\text{Terug: } s = s \cdot B'$$

$$B' = \text{getransponeerde}$$



N-dim basisvectoren

Inproduct-eis basisvectoren:

$$b_n \cdot b_m = 1 \quad \text{indien } n=m$$

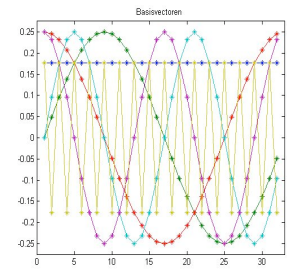
$$b_n \cdot b_m = 0 \quad \text{indien } n \neq m$$

Standaard basis

- Pulsen
- Signaal als gewogen pulsen (PCM = pulse code modulation)

Fourierbasis

- $\cos(k \omega t)$, $k=0:N/2$
- $\sin(k \omega t)$, $k=1:N/2-1$
- Signaal als som van gewogen sinusoiden (golven)



Fourierbasis in N-dimensies

Tijddomein

- pulsen $p_1, p_2, \dots, p_N$

Frequentiedomein

- Golf getal k: aantal perioden per basisvector

$$\cos(\omega k): k=0:N/2$$

$$\sin(\omega k): k=1:N/2-1, \text{ want } k=0, N/2 \text{ geeft } [0 \ 0 \ \dots \ 0]$$

k=0	k=1:N/2-1	k=1:N/2-1	k=N/2
$\cos(\omega k)$	$\cos(\omega k)$	$\sin(\omega k)$	$\cos(\omega N/2)$
$\sqrt{N}$	$\sqrt{N/2}$	$\sqrt{N/2}$	$\sqrt{N}$



Eigenschappen FFT

Efficient te berekenen:
Fast Fourier Transform (FFT)

- In complexe vlak
 $e^{j\omega k} = \cos(\omega k) + j \sin(\omega k)$
 - Reel deel: $\cos(\omega k)$
 - Imaginaire deel: $\sin(\omega k)$
- Voor reëel signaal
 N unieke basisvectoren
 real($e^{j\omega k}$) $k=0:N/2$
 imag($e^{j\omega k}$) $k=1:N/2-1$
- Alle basisvectoren periodiek met periode N

Tijddomein: $x(t)$
Frequentiedomein: $X(k)$

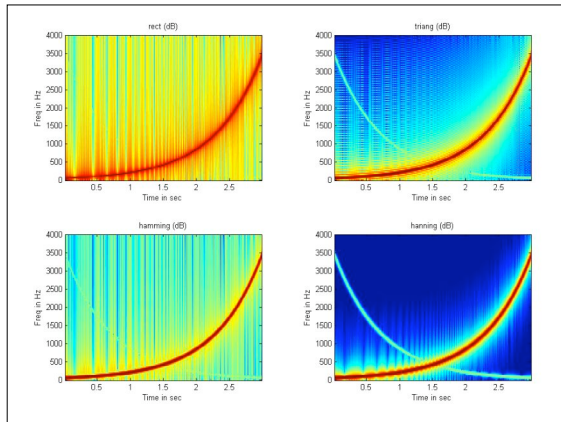
Spectrum $X = \{X_k\}$, $k=0:N/2$
 • Energie x bij frequentie k
 $E(X_k) = X_k \cdot \text{conj}(X_k)$

Spectrogram
 $E(f,t) = \text{tjd-frequentievlak}$

Windowen

- Het isoleren van een stukje signaal
- Niet triviaal !!!!
- want basis is periodiek, maar het signaal niet





Bemonsteren

Bemonsterfrequentie (sample frequency): f_s

- Bepaalt resolutie temporele informatie:
- Bepaalt hoogste frequentie

Nyquist frequentie

- Voor periodiek signaal minimaal top en dal nodig
- $f_{\text{Nyquist}} = f_s / 2$

Telefoonspraak:

$$f_s = 8 \text{ kHz } (dt = 1/f_s = 0.125 \text{ ms})$$

$$\text{Hoogste frequentie in signaal: } f_{\text{Nyquist}} = 4 \text{ kHz}$$

CD kwaliteit:

$$f_s = 44.1 \text{ kHz } (dt = 1/f_s = 0.025 \text{ ms})$$

$$\text{Hoogste frequentie in signaal: } f_{\text{Nyquist}} = 22.05 \text{ kHz}$$



university of
 groningen

faculty of mathematics
 and natural sciences

artificial
 intelligence

auditory
 cognition

26

Vragen?



university of
 groningen

faculty of mathematics
 and natural sciences

artificial
 intelligence

auditory
 cognition

Eenheden FT (tijdsignaal)

Tijddomein

Frequentiedomein

- Sample Periode
 $dt = 1/f_s$
- Duur N-punt window
 $T_{\text{window}} = dt \cdot N$
- Tijd-as
 $T = dt : dt \cdot T_{\text{window}}$

- Samplefrequentie:
 $f_s = 1/dt$
- Basisvector 1
 $f_s = 1/T_{\text{window}} = dt \cdot nfft = f_s/N$
- Hoogste unieke frequentie
 $f_{\text{Nyquist}} = f_s / 2$
- Frequentie-as
 $F = 0 : f_s : f_{\text{Nyquist}}$



university of
 groningen

faculty of mathematics
 and natural sciences

artificial
 intelligence

auditory
 cognition

28



university of
 groningen

faculty of mathematics
 and natural sciences

artificial
 intelligence

auditory
 cognition

29