

# Spraakherkenning en -synthese

David Weenink

Institute of Phonetic Sciences  
University of Amsterdam

First semester 2005

Spraakherkenning  
en -synthese

David Weenink

Administrativa

OS and Software

Inhoud

De golfvorm

Elementaire (basis)  
signalen

De sinus en cosinus  
functies

Het computerpro-  
gramma  
PRAAT

Fourier  
transformatie

Decibel

De spraakketen

Fonetiek en  
fonologie

Spreken

Verstaan

Spraakklanken

Manier van articulatie

Plaats van articulatie

Stemgeving

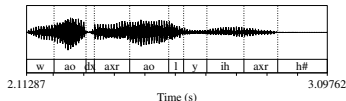
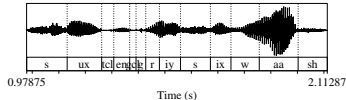
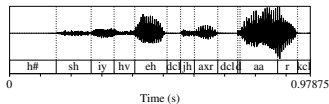
Consonanten/medeklin

- ▶ Colleges: donderdags van 12 – 14
- ▶ Practicum: donderdags 14.30-16.30
- ▶ Wekelijkse opdrachten *Alle opdrachten moeten voldoende voor tentamen*
- ▶ Communication via <http://blackboard.ic.uva.nl/>
  - ▶ Announcements
  - ▶ Mail
  - ▶ Assignments
  - ▶ ...

- ▶ Praat voor analyses: <http://www.praat.org>
- ▶ OS: Linux /Windows/Macintosh

- ▶ Daniel Jurafsky & James H. Martin (2000), *Speech and Language Processing*, Prentice Hall.
- ▶ Kennismaking met (eigen) spraak
- ▶ Spraakanalyses met PRAAT
- ▶ Aspecten van Spraaktechnologie
  - ▶ Synthese
  - ▶ Herkenning
  - ▶ Grote databases
- ▶ Synthese via festival

# TIMIT voorbeeld: sa1 van spreker mjsw0



TIMIT Zin sa1:  
She had your dark suit  
in greasy wash water all  
year

# Decompositie van complexe signalen

Spraak golfvorm is te complex  
Ontbinding via Fourier-transformatie

Spraakherkenning  
en -synthese

David Weenink

Administrativa

OS and Software

Inhoud

**De golfvorm**

Elementaire (basis)  
signalen

De sinus en cosinus  
functies

Het computerpro-  
gramma  
PRAAT

Fourier  
transformatie

Decibel

De spraakketen

Fonetiek en  
fonologie

Spreken

Verstaan

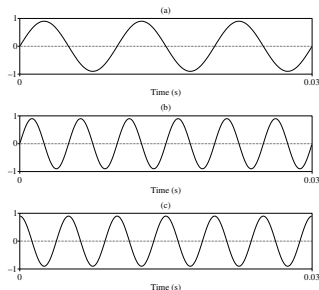
Spraakklanken

Manier van articulatie

Plaats van articulatie

Stemgeving

Consonanten/medeklin



Mathematical functions:

(a)  $0.9 \sin(2\pi 100t)$

(b)  $0.9 \sin(2\pi 200t)$

(c)  $0.9 \sin(2\pi 200t + \pi/2)$

(c)  $0.9 \cos(2\pi 200t)$

## Schaamteloos uit de folder

**flexibel** Interactief & scriptbaar

**portabel** Linux, Macintosh, Windows & Unix versies

**herbruikbaar** Voor Fonetiek, Fonologie, Statistiek

**bruikbaar** Algemeen aanvaarde principes van  
mens-machineinteractie

**onderhoudbaar** Lichtgewicht, goed onderhoudbare code  
(object geïoriënteerd)



- ▶ Scripting
- ▶ History mechanisme
- ▶ Buttons...
- ▶ Dynamisch menu

- ▶ ANSI-C
- ▶ Speciale macros maken lineaire overerving mogelijk
- ▶ Andere macros genereren bij elke klasse automatisch lees, schrijf en kopiëer code
- ▶ Vensteromgeving is Motif met emulaties voor Windows & Mac
- ▶ aangepaste audio

## Objectgeïënteerd ontwerp met klassen

**Algemene klassen** Sound, LongSound, Matrix, PointProcess,  
Strings, TableOfReal, Permutation,  
Distributions, Eigen, SVD, Minimizer, Polygon

**Periodiciteitsanalyse** Pitch, Intensity, Harmonicity

**Spectrale analyses** Spectrum, Spectrogram, Formant,  
LPC, Cepstrum, CC, LFCC, MFCC, Excitation,  
Cochleagram, LTAS

**Labeling** TextGrid, IntervalTier, PointTier

**MDS** (Dis)similarity, Distance, Configuration,  
Procrustus, Confusion, ScalarProduct

**Multivariaat** SSCP, Covariance, Correlation, PCA,  
Discriminant, ClassificationTable,  
ContingencyTable

**Neurale netten** Pattern, Categories, FFNet, ART, ART2A,  
FuzzyART, ARTMAP, CategoryART

**Optimality theory** OTGrammar

## Bruikbaarheid door konsekvente syntax

- ▶ Opschrift op knoppen: Hoofdletter
- ▶ Alléén interactie knoppen hebben "..."
- ▶ Acties zonder nieuw object: *Werkwoord*
- ▶ Acties met nieuw object: "To "
- ▶ Acties in script: opschriften op knop
- ▶ Muisselectie in script: *select*, *plus* en *minus*
- ▶ Scripting: for, while, if, etc.
- ▶ Dynamisch menu:  $\uparrow$ type $\downarrow$  help, Edit, Draw, Query, Modify

Menu: New/Sound

(a) Create Sound... s1 0 0.03 22050  
0.9\*sin(2\*pi\*100\*x)

(b) Create Sound... s2 0 0.03 22050  
0.9\*sin(2\*pi\*200\*x)

(c) Create Sound... s3 0 0.03 22050  
... 0.9\*sin(2\*pi\*200\*x+pi/2)

(c) Create Sound... s4 0 0.03 22050  
0.9\*cos(2\*pi\*200\*x)

# Decompositie: Fourier transform

## Nieuwe basis

1. sum of sines and cosines

$$s(t) = \sum_{k=0} (a_k \sin(2\pi k f_0 t) + b_k \cos(2\pi k f_0 t))$$

2. sum of sines with phases

$$s(t) = \sum_{k=0} (c_k \sin(2\pi k f_0 t + \phi_k)) \text{ (Immers:}$$
$$a \sin \alpha + b \cos \alpha = c \sin(\alpha + \phi),$$

met  $\alpha$  en  $\phi$  functies van  $a$  and  $b$ )

## In PRAAT

select Sound s1

To Spectrum... no<sup>1</sup>

Edit

---

<sup>1</sup>Option yes: perform Fast Fourier Transform.

## Decibel

$$\text{dB's: } 10 \log \frac{P(\text{ower})}{P_{\text{ref}}}$$

Omdat  $P \sim \text{Amplitude}^2$

$$\text{dB's: } 20 \log \frac{A(\text{mplitude})}{A_{\text{ref}}}$$



Van idee naar geluid naar perceptie naar idee naar geluid...

- ▶ Fonetiek:
  - ▶ fysieke proces van spreken en verstaan  
productie, signaaleigenschappen, verschillen...
  - ▶ analyse spraaksignaal
- ▶ Fonologie: klanken onderdeel van systeem
  - ▶ welke zijn betekenisonderscheidend
  - ▶ hoe is het klanksysteem?
  - ▶ welke combinaties mogen?

Fonetisch: [a]

Fonologisch: /a/

# De spraakorganen

Spraakherkenning  
en -synthese

David Weenink

Administrativa

OS and Software

Inhoud

De golfvorm

Elementaire (basis)  
signalen

De sinus en cosinus  
functies

Het computerpro-  
gramma  
PRAAT

Fourier  
transformatie

Decibel

De spraakketen

Fonetiek en  
fonologie

Spreken

Verstaan

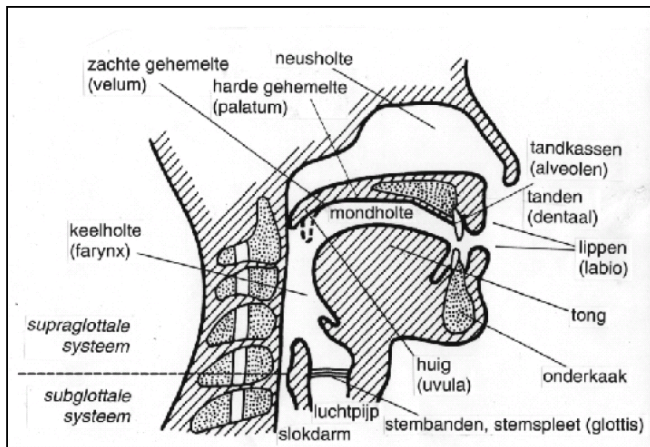
Spraakklanken

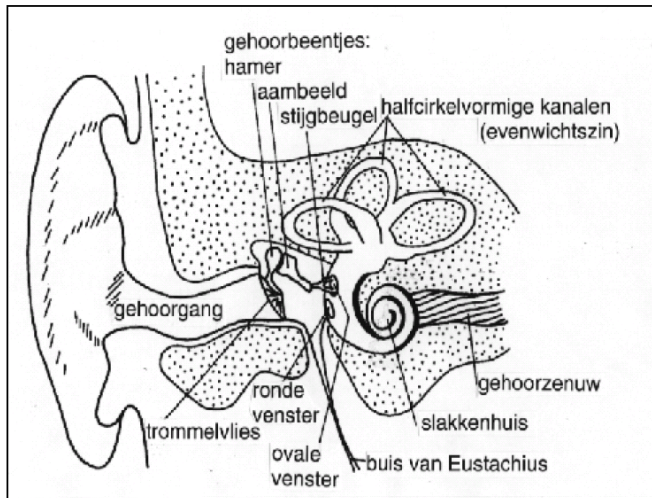
Manier van articulatie

Plaats van articulatie

Stemgeving

Consonanten/medeklin





## Twee categoriën

- ▶ vocalen /klinkers  
nauwelijks vernauwing in mond-keel kanaal
- ▶ consonanten / medeklinkers  
Wel vernauwing

## Verder onderscheid op basis van

- ▶ Manier van articulatie
- ▶ plaats van articulatie
- ▶ stemgeving

- ▶ plosieven (plofklanken): p, t, k  
volledige afsluiting met overdruk, release
- ▶ fricatieven (wrijfklanken): f, s  
bijna volledige vernauwing
- ▶ liquidae (vloeiklanken): r, l  
lucht langs zijkant tong
- ▶ nasalen (neusklanken): m, n  
lucht door de neus
- ▶ halfklinkers (glijklanken): w, j  
nauwelijks vernauwing

# Plaats van articulatie

p, t, k zijn verschillend

- ▶ labiale klanken: b  
bij de bovenlip
- ▶ dentale klanken: d  
bij de boventanden
- ▶ alveolaire klanken: s  
tandkas achter de boventanden
- ▶ palatale klanken: j  
harde gehemelte
- ▶ velaire klanken: k  
zachte gehemelte
- ▶ uvulaire klanken: huig-r  
bij de huig
- ▶ glottale klanken: h  
bij de stemspleet

dentaal+alveolair+palataal = coronale klanken

velair+uvulair+glottaal = dorsale klanken

Spraakherkenning  
en -synthese

David Weenink

Administrativa

OS and Software

Inhoud

De golfvorm

Elementaire (basis)  
signalen

De sinus en cosinus  
functies

Het computerpro-  
gramma  
PRAAT

Fourier  
transformatie

Decibel

De spraakketen

Fonetiek en  
fonologie

Spreken

Verstaan

Spraakklanken

Manier van articulatie

**Plaats van articulatie**

Stemgeving

Consonanten/medeklin

- ▶ stemhebbend: b, d, g
- ▶ stemloos: p, t, k



# Manier en plaats van articulatie

<i>plaats</i> →	labiaal		coronaal				dorsaal	
			dentaal, alveolair		palataal			
<i>stemgeving</i> →	-stem	+stem	-stem	+stem	-stem	+stem	-stem	+stem
<i>manier</i> ↓								
plofklank	<b>p</b> <i>paal</i>	<b>b</b> <i>baal</i>	<b>t</b> <i>taal</i>	<b>d</b> <i>dop</i>	<b>tʃ</b> <i>checken(E)</i> <b>tʃ</b> <i>tjalk</i>	<b>dʒ</b> <i>jeep(E)</i> <b>dʒ</b> <i>Jakarta</i>	<b>k</b> <i>kok</i> <b>ʔ</b> <i>aap</i>	<b>g</b> <i>goal(E)</i>
fricatief	<b>f</b> <i>fiets</i>	<b>v</b> <i>vies</i>	<b>s</b> <i>sier</i>	<b>z</b> <i>zier</i>	<b>ʃ</b> <i>sjaal</i>	<b>ʒ</b> <i>rouge(F)</i>	<b>x</b> <i>acht</i> <b>h</b> <i>huis</i>	<b>ɣ</b> <i>gele</i>
nasaal		<b>m</b> <i>maar</i>		<b>n</b> <i>maar</i>		<b>ɲ</b> <i>oranje</i>		<b>ŋ</b> <i>ring</i>
liquida				<b>l</b> <i>leuk</i> <b>r</b> <i>reuk</i>				<b>ʀ</b> <i>reuk</i>
halfklinker		<b>w</b> <i>week</i>				<b>j</b> <i>jeuk</i>		

## Articulatieplaats

- ▶ positie van de tongrug
  - ▶ voor-achter: /i/, /u/
  - ▶ hoog-laag of gesloten-open: /u/ - /a/
- ▶ lippen: /i/, /y/

<i>plaats van vernauwing</i> →	voor		midden			achter	
<i>lipvorm</i> →	gespreid	ongerond	gerond	ongerond	gerond	ongerond	
<i>mate van vernauwing</i> ↓							
hoog	i kies		y fuut			u toen	
midden		i e: kɪp keet	y ø: hut peuk	ə de	ɔ o: top pook		
laag		ɛ tel		a: taak		ɑ tak	

## Manieren

- ▶ nabootsing van menselijke articulatie
- ▶ recorder: opslag van alle zinnen
- ▶ opbouw uit kleinere stukjes
  - ▶ difoonsynthese extra = [#ε][εk][ks][st][ra:][a:#]  
beregeling van luidheid, foneemduur (beklemtoonde  
wordt langer) en toonhoogte (vraagzin)
- ▶ totaal via synthesesregels