

Språkteknologi – En Introduktion

Robert Eklund

robert@roberteklund.info
http://roberteklund.info



Universitetslektor (Fonetik) vid
Institutionen för Kultur och Kommunikation, Linköpings Universitet
Linköping, Sverige

&
Affilierad Forskare (Neurokognition) vid

Avdelningen för Neurovetenskap / Karolinska Institutet
Stockholm, Sverige



&
Docent (Datorlingvistik) vid



Avdelningen för Datavetenskap, Linköpings Universitet
Linköping, Sverige

30 september 2012

Denna presentation

- Vad är språk?
- Vad är språkteknologi?
- Vad är språkteknologins delområden
- Varför språkteknologi?
- Historik
- Vartåt barkar det hela?

< BILD HÄR >

Lt Commander Data i Star Trek

2

Olika områden

- Taligenkänning
- Talsyntes
- Text-till-Tal-System
- Talarverifiering (-identifiering)
- Maskinöversättning
- Dialogsystem
- Ansiktsanimering
- Multimodala system
- Androider
- "Neuroteknologi"

3

1. Introduktion

4

Introduktion (1)

- Vi vet inte när talet utvecklades
- Skriftsystem 5000 år gamla (Mesopotamien, Egypten, Kina)
- Fullt utvecklade språk, tal mycket äldre
- Mänsklig hjärna och talorgan är anpassade för talproduktion
- Schimpanser *kan* inte producera talljud
- Hoover the Talking Seal
 - * 1971, Maine
 - † 1985, Boston



5

Introduktion (2)

- Det finns 5000–8000 språk
- Ojämnt distribuerade
- Engelska 2:a / svenska 85:e plats
- Teknologisk utveckling → Tala med datorer

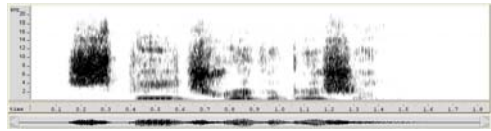
6

2. Taligenkänning

7

Taligenkänning (1)

- Datorer som "förstår" tal
- Spektrografen uppfanns 1946 → "läsa" tal



- Första systemet beskrivet i [Davis, Biddulph & Balashek \(1952\)](#)

8

Taligenkänning (2)

- Talets variabilitet mycket större än man trott
- Denna variabilitet är (fortfarande) problemet
- Tal varierar (bl a) med avseende på...
 - Kön
 - Alder
 - Dialekt
 - Sociolekt
 - Individ (mellan/inom)
 - Talhastighet/reduktioner ("har du en bil" → "harumbil")
 - Disfluenser ("eh", "öh" etc)

9

Taligenkänning (3)

- Taligenkännare "tränas"
- Skapar meningar som täcker/beskriver det aktuella språket (fonetiskt/lingvistiskt)
- Talare av språket spelar in meningarna
- Talarna bör vara balanserade över kön, ålder, dialekt, sociolekt och så vidare
- Datorn jämför inspelningarna med transkriptioner
- Datorn lär sig variabiliteten
- Men vad, exakt, ska man träna på?

10

Taligenkänning (4)

- Lingvistisk term: fonem
- Definition: "Minsta betydelseändrande enhet"
- 1. b/p ändrar betydelsen i svenska, men inte i finska
Ex: *bil* vs *pil*
- 2. v/w ändrar betydelsen i engelska, men inte i svenska
Ex: *vie* vs *why*
- Antalet fonem i olika språk varierar mellan 11 (Rotokas) och 141 (IXÚ)
- Engelska och svenska: ca 45

11

Taligenkänning (5)

- Två skilda termer:
 - Tal-igenkänning*
... känner igen *vad* som sägs, oavsett *vem* som säger det
 - Talar-igenkänning (-identifiering/-verifiering)*
... känner igen *vem* som talar, oavsett *vad* som sägs
- Barn är svårare att känna igen än kvinnor, som är svårare att känna igen än män
- Akustiska orsaker: Ju högre grundton, desto större "sökrymd"

12

3. Talsyntes

13

Talsyntes

- Datorer som pratar/talar
- Tre äkta/genuina syntesformer:
 1. Artikulatorisk syntes
 2. Formant-syntes
 3. Konkaterings-syntes
- Pseudosyntes:
”Burkat tal”

14

Burkat tal (“canned speech”)

- Man spelar in en ”bärfras”:
Saldot på kontot är _____ kronor.
- Fyller i den tomma luckan med förinspelat material
- Varje nytt yttrande kräver en ny inspelning, helst med samma person/röst (som kanske bytt jobb)
- Delvis bra ljudkvalitet
- Låg flexibilitet

15

Artikulatorisk Syntes (1)

- Försöker imitera beteendet hos människans talorgan
- Använder virtuella...
 - lungor
 - stämläppar
 - kinder
 - tunga
 - gom
 - luftströmmar
 - läppar
 - tänder
- ... med mera och så vidare et cetera

16

Artikulatorisk Syntes (2)

- Urban Hjärne (1641–1724)
- Använde avhuggna huvuden
- Bälgar pumpade luft genom strupen
- Trådar styrde tunga, läppar, kinder
- Inte så bra resultat
- Etiska problem



17

Artikulatorisk Syntes (3)

- Alexander Graham Bell (1847–1922)
- Använde sin (levande!) hunds strupe
- Hunden ylade, Bell klämde



- Kunde producera få ljud
- Inte så bra resultat
- Etiska problem

18

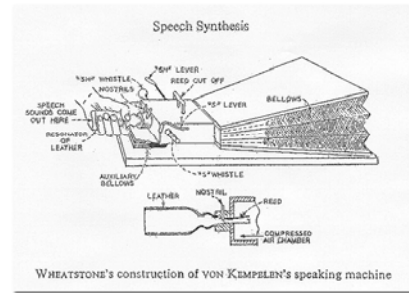
Artikulatorisk Syntes (4)

- Wolfgang von Kempelen (1734–1804)
- Skapade en talmaskin (1791)
- Mekanisk automat
- Fungerade!!



19

Artikulatorisk Syntes (5)



Tack till Mária Gósy, Kempelen Farkas Speech Research Laboratory, Budapest

20

Artikulatorisk Syntes (6)

- Hade tidigare skapat en schackspelande automat (1770)
- "Turken"
- Människa gömd inuti maskinen (dvärg?)
- Dålig "cred"
- Tidigt exempel på "artificiell intelligens"
- Jämför Lt Commander Data (Star Trek)



21

Artikulatorisk Syntes (7)

- Ur Scientific American, 1901

"Dr. Marage has constructed an apparatus, using the plastic substance employed by dentists, to reproduce the interior of a person's mouth while pronouncing the different vowels."



22

Artikulatorisk Syntes (8)

- Hideyuki Sawada, Kagawa University

< FILM HÄR >

http://www.eng.kagawa-u.ac.jp/~sawada/index_e.html

23

Artikulatorisk Syntes (9)

- Svår metod
- Den mest intressanta ur forskningsperspektiv
- Inte så intressant kommersiellt (hittills)
- Om det lyckas kan datorn göra allt som människor kan

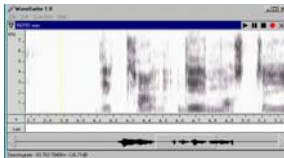
24

Formantsyntes (1)

- Imiterar talets akustiska karakteristika

- **Formanter**

= frekvensband med hög amplitud i talsignalen



- Tonhöjd = melodisk kontur = Formant 0 = F0
- Genererar tonpuls i datorn
- Filtrerar tonen elektroniskt tills det låter som tal

25

Formantsyntes (2)

- Tidig synthesizer:

The Voder (1939)

- Homer Dudley



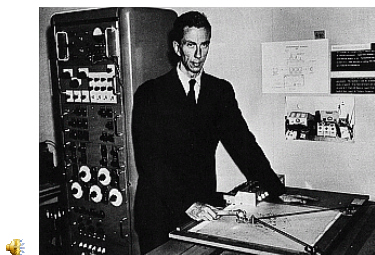
26

Formantsyntes (3)

- Tidig svensk synthesizer (som talar engelska):

Ove (1953)

- Gunnar Fant



27

Konkateneringssyntes (1)

- "Klipp-och-klistrings"-syntes

- Spelar in riktigt mänskligt tal

- Klipper ut ljudbitar

1. Fonetem: Omöjligt! Inte samma [k] i "kille" och "kalle"
2. Difoner: /ki/ka/ku/ko osv
3. Polyfoner: spri-/orvi-olmskt osv
4. Större enheter: *Unit selection*. Godtyckligt stora enheter

- Klistrar ihop dem i ny ordning

28

Konkateneringssyntes (2)

- Inspelning:

Ø	<u>ta</u>	IGEN
Säga	<u>t</u> at	IGEN
Säga	ta <u>l</u>	IGEN
Säg	als <u>a</u>	IGEN
Säga	spr <u>å</u> t	IGEN
Säga	tå <u>k</u>	IGEN
SÄGA	ta <u>k</u>	Ø

- Syntes: t + ta + al + ls + språ + åk + k

29

Konkateneringssyntes (3)

- Antalet polyfoner varierar mellan språk

- *Fonotax*: Hur fonem tillåts kombineras

- Svenska tillåter många kombinationer

- Typisk stavelsestruktur: CV- eller V-stavelser

- Svenskt maximum: CCCVCCCCCCCC

Ex: CCV ... VCCCCC
"... ordet stockholmskts ... uttal."

- Mänsklig talkvalitet

- Kommersiellt intressant, speciellt *unit selection*

30

Syntesexempel (1)

1. Inspelning: Kvinnoröst 
2. Syntes: Kvinnoröst 
3. Inspelning: Mansröst 
4. Syntes: Mansröst 
5. Syntes: Manspolyfoner / Kvinnlig grundton 
6. Syntes: Kvinnopolyfoner / Manlig grundton 

31

Syntesexempel (2)

7. Syntes
7.1 Formantsyntes **Namnupplysningen** 
Ove (KTH)
7.2 Polyfonsyntes *Prophon* (Telia 1995) 
7.3 Polyfonsyntes *Prophon* (Telia 1995) 
Telefonkvalitet: 300–3400 Hz
8. Engelsk syntes **Flygbokningsmeningar** 
8.1 Difonsyntes *TrueTalk*
9. Fransk syntes **Flygbokningsmeningar** 
9.1 Difonsyntes *CNETVox*

32

4. Text-till-Tal-System

33

Text-till-Tal-System (1)

- Datorer som läser ut text
- Kan använda vilken talsyntesmetod som helst
- Det finns ett otal andra problem
- Allt syns inte i text
- Ett exempel: Partikelverb
Bo stötte på Lena vs *Bo stötte på Lena*
- Annat problem: Uttal av främmande namn/ord
Margaret Thatcher vs *Margaret Tåtser*

34

Text-till-tal-system (2)

- Språk har olika förhållande mellan skrift/ljud:
 - Finska Nästan perfekt
 - Turkiska Nästan perfekt
 - Svenska Ganska dåligt
 - Franska Ganska bra text-till-tal (men hemskt tal-till-text!)
 - Engelska En katastrof!!!! (Charivarius: *The Chaos*)
<http://pages.cpsc.ucalgary.ca/~hill/papers/conc/thechaos.htm>
 - Mandarin ... och andra språk med ikonografiska skriftspråk.
- Intonation över hela satser, och från sats till sats
- Dessa – och många liknande – problem måste lösas innan talsyntesen kan säga något!

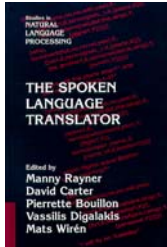
35

5. Maskinöversättning

36

Maskinöversättning (1)

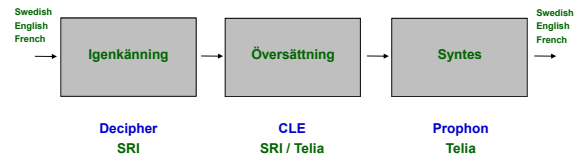
- Automatisk översättning mellan språk
- Text-till-Text eller Tal-till-Tal
- Project: Spoken Language Translator
 - Telia Research
 - SRI International, Menlo Park, California
 - SRI International, Cambridge, UK
- 1993–1999
- Air Travel Information Service (ATIS)
- Engelska/Svenska/Franska
- Bok på Cambridge University Press



37

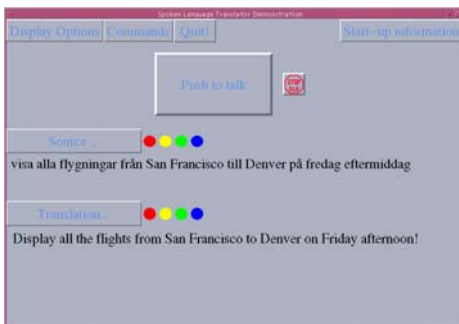
Maskinöversättning (2)

Spoken Language Translator



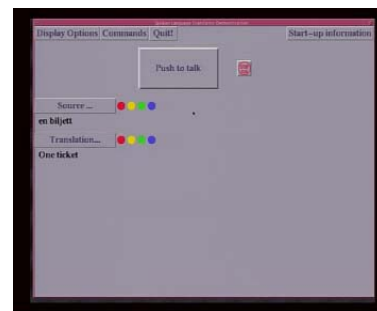
38

Maskinöversättning (3)



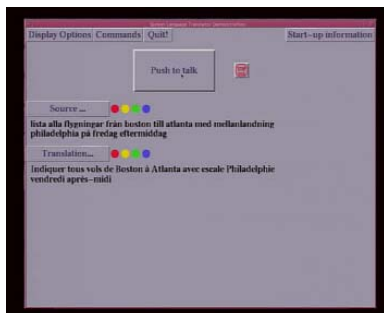
39

Maskinöversättning (4)



40

Maskinöversättning (5)



41

6. Dialogsystem

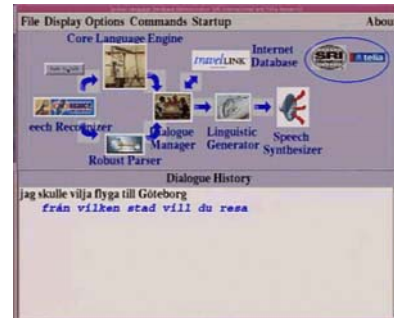
42

Dialogsystem (1)

- System som både lyssnar och talar
- Kommunikation med t ex databaser
- Kräver annat slags lingvistisk kunskap
- "Konversationsgrammatiker"
- "Dialoghantering"
- Kommerciellt mycket intressant

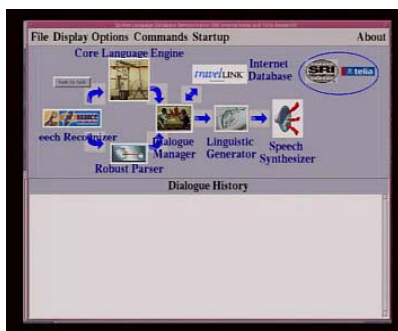
43

Dialogsystem (2)



44

Dialogsystem (3)



45

7. Ansiktsanimering

46

Ansiktsanimering (1)

- Talsyntes med ansikte
- Förbättrar förståelsen av talsyntes
- McGurk-effekten
- Alla "hör" med ögonen (inte bara hörselskadade)
- Originalreferens:

Harry McGurk & John MacDonald. 1976.
Hearing lips and seeing voices,
Nature, vol. 264, pp. 746-748

47

Ansiktsanimering (2): McGurk

< FILM HÄR >

<http://www.youtube.com/watch?v=aFPtc8BVdJK>

48

Ansiktsanimering (3)

- Finns olika metoder
- Telia Research: unik metod
- Samma princip som talsyntesen
- Kommer inte ihåg? OK...

49

Ansiktsanimering (4)

- Inspelning:

Ø	<u>ta</u>	IGEN
Säga	<u>ta</u> t	IGEN
Säga	ta <u>l</u>	IGEN
Säg	als <u>a</u>	IGEN
Säga	spr <u>å</u> t	IGEN
Säga	tå <u>k</u>	IGEN
SÄGA	tak <u> </u>	Ø
 - Syntes: t + ta + al + ls + språ + åk + k
- ... lägg till ansiktsrörelser!

50

Ansiktsanimering (5)

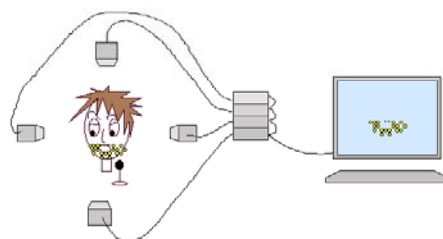
- Spela in ansiktsrörelser tillsammans med ljudet
- Reflektorer i ansiktet på talaren
- 24 reflektorer runt munnen (och näsan)
- Glasögon för att normalisera för huvudrörelser



51

Ansiktsanimering (6)

- Inspelning:



52

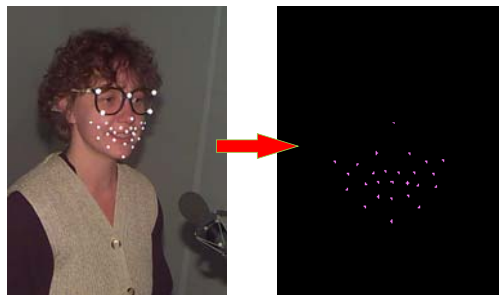
Ansiktsanimering (7)

- Labbet:



53

Ansiktsanimering (8)

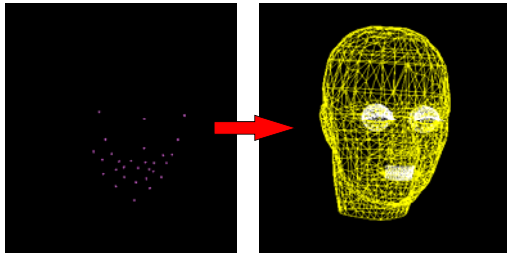


Reflektorrörelser lagras som...

... 3D-rörelser i datorn

54

Ansiktsanimering (9)



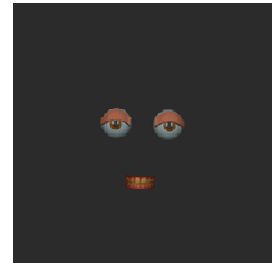
Reflektorerna fästs på...

... en trådmodell av ett huvud

55

Ansiktsanimering (10)

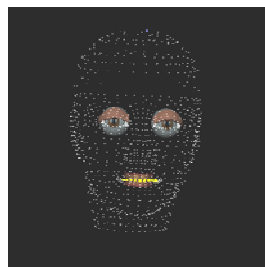
- Lägg till ögon, tänder...
- Datorgrafik



56

Ansiktsanimering (11)

- Huvudform:



57

Ansiktsanimering (12)

- Lägg till textur

Vattenkammat hår...



... Göm reflektorerna

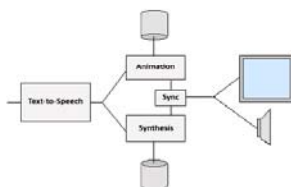


Foton av riktigt ansikte...

58

Ansiktsanimering (13)

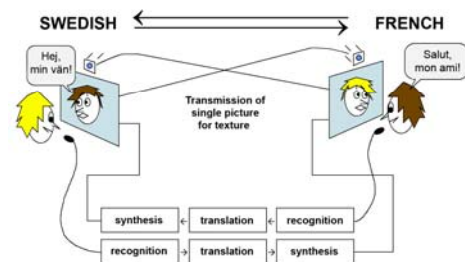
- Systemet:



59

Ansiktsanimering (14)

- Vision:



60

Ansiktsanimering (15)



61

8. Multimodala gränssnitt

62

Multimodala gränssnitt (1)

- Kombination av olika modaliteter
 - Tal
 - Mus
 - Gester
 - Ansiktsuttryck
 - Känsel
 - Lukter
 - ... osv
- Både input/användare och output/dator

63

Multimodala gränssnitt (2)

- Hur interagerar människor med datorer?
- **Reeves & Nass (1996):**
The Media Equation
- "It's Only A Movie"-fenomenet
- **Nass, Kim & Lee (1998):**
When Your Face is the Interface: An experimental comparison of interacting with one's own face or someone else's face
- **Jönsson & Dahlbäck (1988):**
Talking to your computer is not like talking to your best friend

64

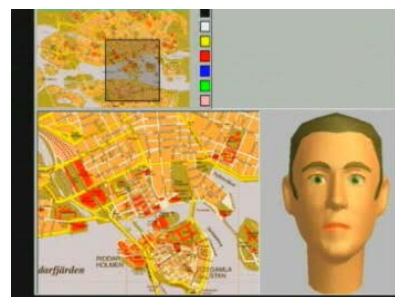
Multimodala gränssnitt (3)

- AdApt
- Telia och KTH
- Lägenhetsinformation



65

Multimodala gränssnitt (4)



66

Multimodala gränssnitt (5)

- Waseda University (Japan): **Humanoid**
- Konversationsrobot
- Integrerar:
 - Taligenkänning
 - Talsyntes
 - Dialoggrammatik
 - Datorseende
 - "Gaze Tracking" (ser vem en person ser på)
 - Fungerande robotik

67

Multimodala gränssnitt (6): Humanoid

< FILM HÄR >

<http://www.humanoid.waseda.ac.jp/>

68

Multimodala gränssnitt (7): Aiko

< FILM HÄR >

<http://www.aikoproject.jp/>
<http://www.youtube.com/watch?v=lg7qmdOq4s&feature=related>

69

9. "NeuroTechnology"

70

"NeuroTechnology"

- Inte språkteknologi i strikt bemärkelse
- Dock är språk nästan alltid inblandat
- Kort (!) sammanfattning av några områden

71

Brain-Reading

- Via "neuroimaging" läser man hjärnaktivitet
- EEG: Hjälms/nät av elektroder registrerar elektriska aktivitet
- fMRI: Mäter syreförbrukning (antas vara aktivitet)
- Via signaler har man lyckats se:
 - ... vad folk tittar på för bild (ansikte, katt, stol etc) (Haxby et al., 2001)
 - ... vilken bild av 1000 bilder som en person tittar på (Kay et al., 2008)
 - ... vilken scen ur Sergio Leones film *Den gode, den onde, den fula* en person just tittar på (Hasson et al., 2004)

< BILD HÄR >

72

Brain-Computer Interfaces (BCI)

- Gränssnitt direkt mellan hjärna och dator
- Hjälms eller nät av elektroder registrerar hjärnans elektriska aktivitet och använder den för att styra dator
- Användningsområde:
"locked-in"-patienter (kvadripleger)
- "Neural prosthetics"
(Andersen et al., 2004)

< BILD HÄR >

<http://computer.howstuffworks.com>

73

Lögn detektion

- Normala lögn detektorer inte pålitliga/tillåtna i domstolar
- Hypotes: olika hjärnaktivitet vid sanning och lögn
- Metod: via EEG eller fMRI direkt
kolla om en viss hjärna producerar en sanning eller en lögn
- Hett område (duh!)

< BILD HÄR >

<http://singularityhub.com/2010/05/06/another-attempt-to-use-fmri-lie-detector-in-us-court-falls-in-brooklyn-more-on-the-way/>

- "Not there yet"

74

10. Sammanfattning

75

Sammanfattning

- Mängd detaljproblem att lösa
- Allmän teknologiutveckling ger nya utmaningar
- Målbilden är minst sagt ambitiös!
- Flera tänkta användningsområden:
 - Personer med handikapp
 - Sällskap
 - Nyhetsankare (?)
- Språk spelar en central roll i skapandet av androider
- ... och glöm inte "Turken"!

76

Sammanfattning (2)

- Så, hur långt har vi kommit?
- Ja, jämför...

< BILD HÄR >

< BILD HÄR >

Lt Commander Data

Aiko

77

Bonus

... och slutligen en kommentar, "as seen on TV" ...

78

TV3 Direkt, 3 november 1998



79

Tack för uppmärksamheten!

Kontakta mig gärna:

robert@roberteklund.info

Hemsida:

<http://roberteklund.info>

80