



Lektor Børge Lindberg

Speech & Multimedia Communication Division

Afdeling for Kommunikationsteknologi

Aalborg Universitet

lindberg@kom.auc.dk



Emner

1. Hvad er talegenkendelse
2. Teknikker
3. Anvendelser
4. Produkter
5. Forskningsemner og -projekter
6. Danske forhold og det danske sprog



Ordgrænser: dedændæ'h

Grammatik : Det [er] den der

Udtale : Andre lydverdier 'e/e' 'e/æ/' 'r'

Melodi: Spørgsmål '?'

Tryk: Det er den der / Det er den der

Enslydende: vår – hvor

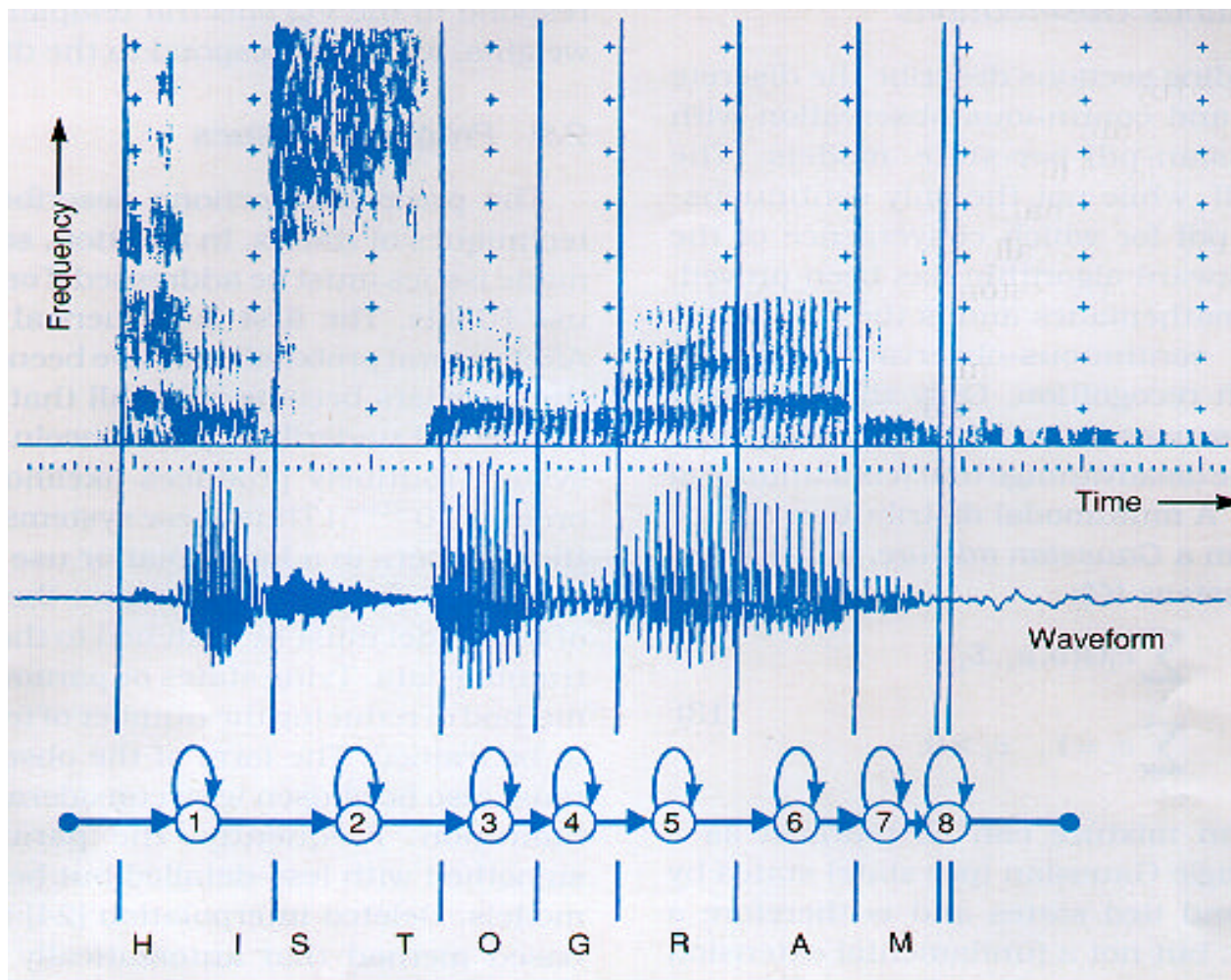
Forskelligt lydende: da'r – de'r (koen der lo, dér er den)

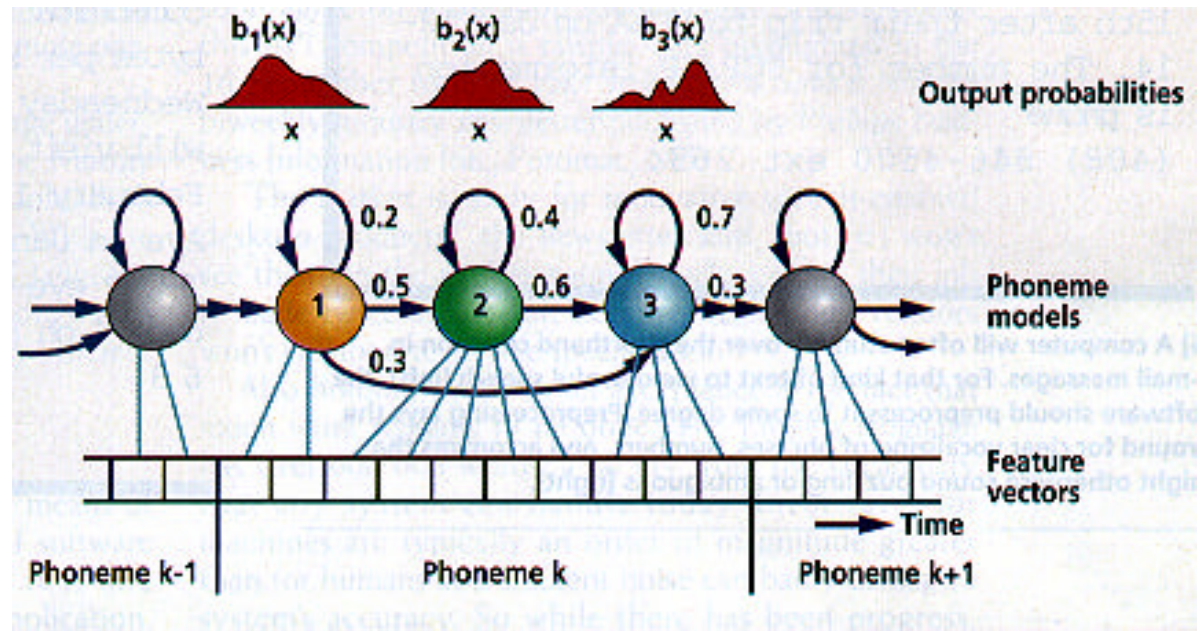
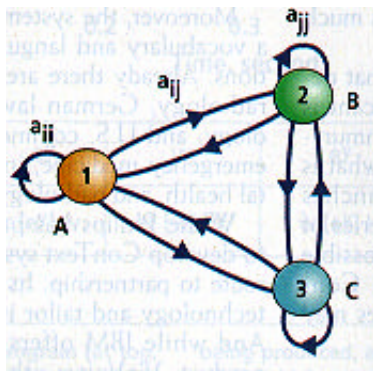
Udtaleforskelle : Dialekt, Mand/kvinde, Ældre/ung/barn,
Modersmål/fremmedsprog, Individuelle forskelle

Dansk er specielt – men ikke specielt svært

Overgås dog kun af engelsk i ikke-lydrethed (synteseproblem)

Princip (i de fleste talegenkendere)





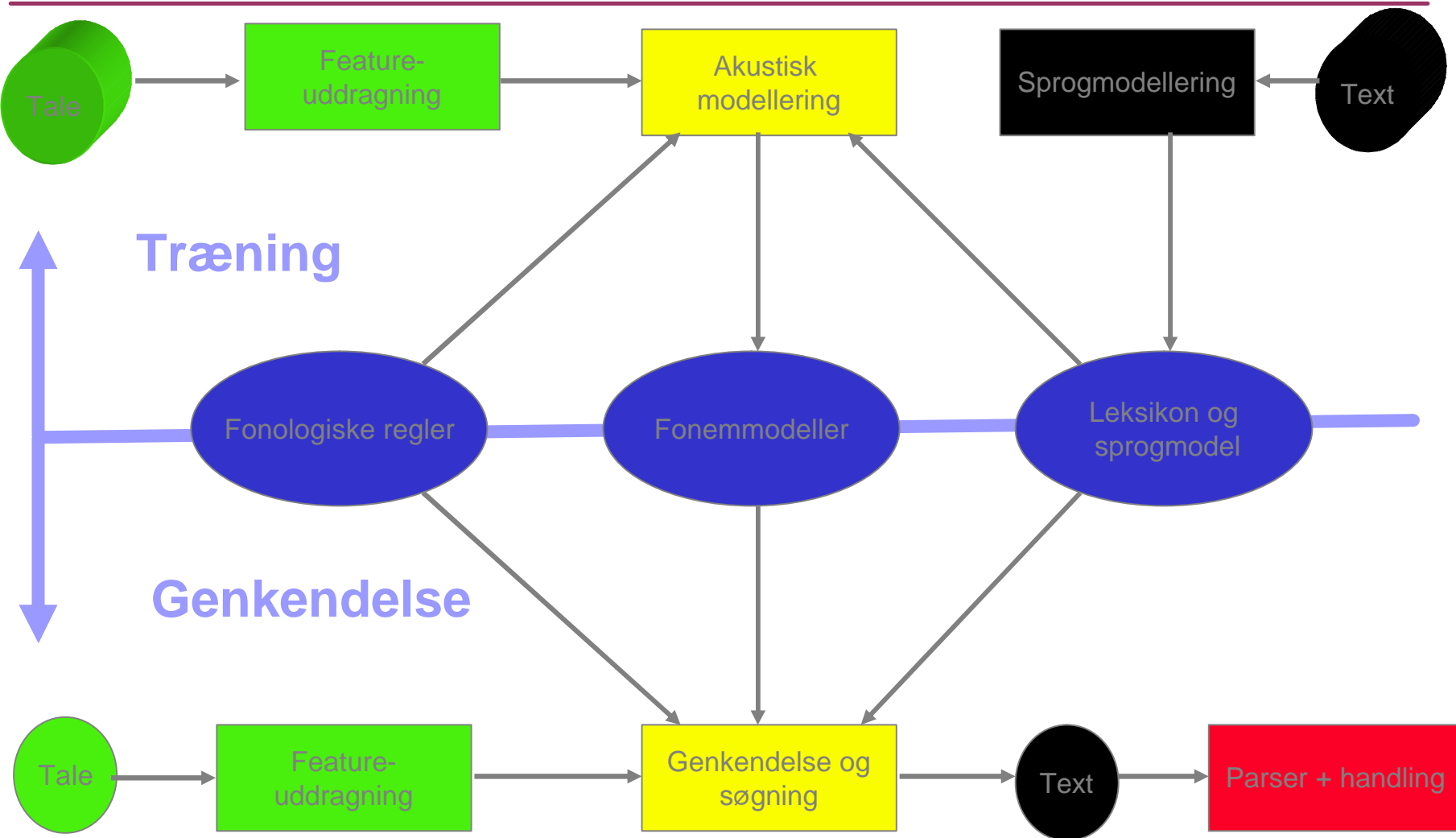
© IEEE Spectrum December 1997



Hardware !

Erkendelse: Corpus baseret teknologi

- **akustiske databaser, sproglige databaser (tekst), sproglige databaser (vendinger), dialog databaser (hyppighed og betydning af udtryk)**
- **Hjulpet af udvikling i medier (CD-ROM, DVD, Harddisk : TB !)**
- **Linguistic Data Consortium (LDC)**
www ldc upenn edu
- **European Language Resources Association (ELRA)**
<http://www icp grenet fr/ELRA/home.html>





Uden trykknapper

- Ikke alle telefoner og centraler kan bruge trykknapper

Håndfri

- Arbejdssituationer, fx bilkørere, flymekanikere, køkken.

Naturligt og evt. nødvendigt

- Mindre indlæring end med tastatur og mus
- <http://www.spok.dk>
- AMI, Bispebjerg, 4 store danske virksomheder

Funktionshæmmede (handicappede)

- Kommunikation eller styring af apparater/udstyr
- Personer med dysartri (?), motorisk hæmmede, kommunikation med døvblevne, blinde, dyslektikere



Information og bestillinger

- Tog, banker, aviser, call centre

Diktering

- Forskellige domæner : alm. breve, jura, medicin

Typiske telefonitjenester:

- Kundevejledning (vejledning og svar på typiske spørgsmål)
- Omstilling (sig navnet på den, du vil tale med)
- Bankoplysning (saldi, aktiekurser, info om overførsler)
- Trafikinformation (fly, tog, færges)
- Booking
- Voice portaler (ring til ét nummer (“Sonomemo”) og få oplæst e-mails, kalender, telefonsvarer – talestyret)



- 72 porte (parallele opkald)
- 2,3 mio. opkald pr. år – 32.000 pr. port
- 3.300 \$ pr. port pr. år (i alt 950.000 \$ over 4 år)
- Gennemsnitlig længde pr. opkald : 3 min

Key Comparisons between Agents and IVR

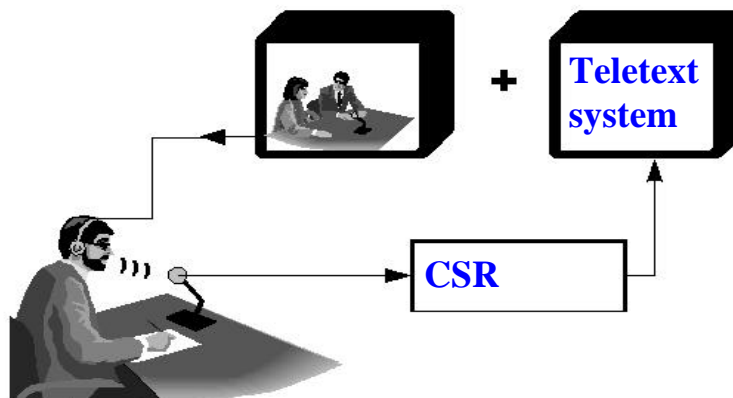
Metric	Live Agent	Speech-enabled IVR Port
Calls per Year	28,350	32,000
Cost per Year	\$30,000	\$3,300
Cost per Call	\$1.06	\$0.10



TV studio



Operator room





Velotype system (baseret på stenografi) kræver to operatører - en til opsummering, en til stenografi. Bruges hos BBC, UK og i Holland.

I Sverige bruges kun én operatør - problematisk.

Talegenkendelse bruges allerede:

- I Tyskland: ved ZDF, Mainz bruger IBM ViaVoice Pro, <http://www-4.ibm.com/software/speech/de> , siden juni 2001
- I Spanien : Har brugt IBM ViaVoice Pro de sidste 5 år – virker godt
Ofte semi-live (8-9 sec. delay). Max. 80-90 ord pr. minut.
CNN : Ofte 200 ord pr. Minut, JFK : 300 words pr. minut
- I UK, BBC : Overvejer brugen af ViaVoice Pro
- Danmark : Kræver først og fremmest tilstedeværelsen af generel dansk talegenkendelse – desuden speciel brugergrænseflade ("realtime editing")



Rose, ICSLP'02

Wireless/Mobile devices

Tablet PC



Fujitsu
Stylistic LT C-500

PDA



Sony
CLIE PEG-N710

PDA / Mobile Phone



Handspring
Treo 270

Mobile Phone



Motorola
M70

- Non-PC Internet connections:
- 1998: 4 %
- 2001: 40%
- 2003: More than PC's

- Lille eller intet keyboard
- Input/Output modality selection
- Mindre beregningskapacitet end PC'er
- Client-server arkitektur



SAS trafik information

- Information om flyafgange, ankomst, m.m.
- Ej billetbestilling
- Ring: 70102000 - 22
- Flydende tale, systemstyret dialog, www.nuance.com

TDC

- Telefonservice, udland:
- landekoder, priser, klokkeslet – god word-spotting
- Ring: 80604055
- Nummeroplysning: Ring: 1811
- By, gade, navn indtales



ATP

- Generel vejledning
- Ring 48204910 – Tryk 3
- Prolog Development Center (www.pdc.dk/speechlogic)

Philips toginformation

- Information om tyske intercitytog
- Brugeren fortæller hvorfra, hvortil, dag, og enten cirka afgangstid eller ankomst, og systemet giver: præcise afgange/ankomster samt mellemskift.
- Ring: +49 241 60 40 20, den forstår ”dansk-tysk”!
- Flydende tale, systemstyret dialog, 2000 ord, heraf 1100 stationer



Ordforråd

- Robuste over telefon 2000+ ord
- Dikteresystemer 50.000+ ord

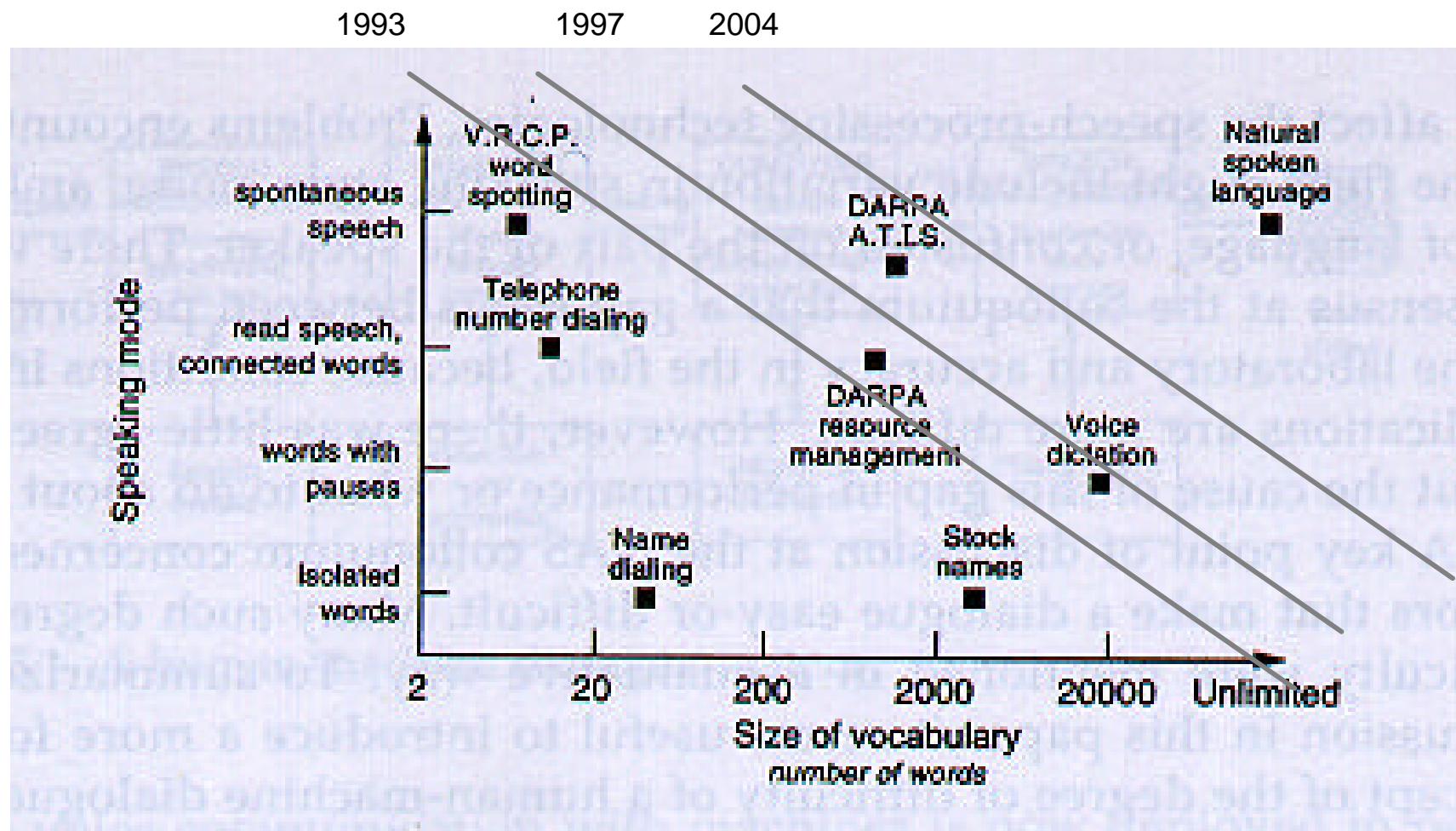
Fejlgenkendelser

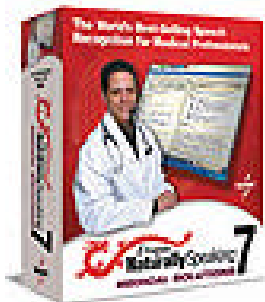
- Systemet har svært ved at ”erkende egne fejl”

Grammatik

- Benægtelser
- Sammensatte sætninger

Udfordringer – opgavens karakter



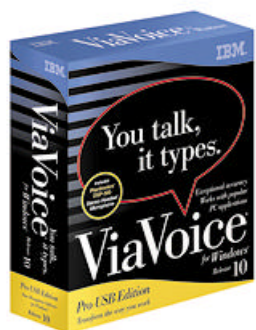


- Dragon NaturallySpeaking, Pro, Medical, Legal eller Preferred (\$ 679, 979, 979, 199) (Scansoft, Command & control)
- IBM ViaVoice (til Windows) \$ 190
- Philips SpeechMagic Patologi/Røntgen \$5.000
- CyberTranscriber (Speech Machines) \$?

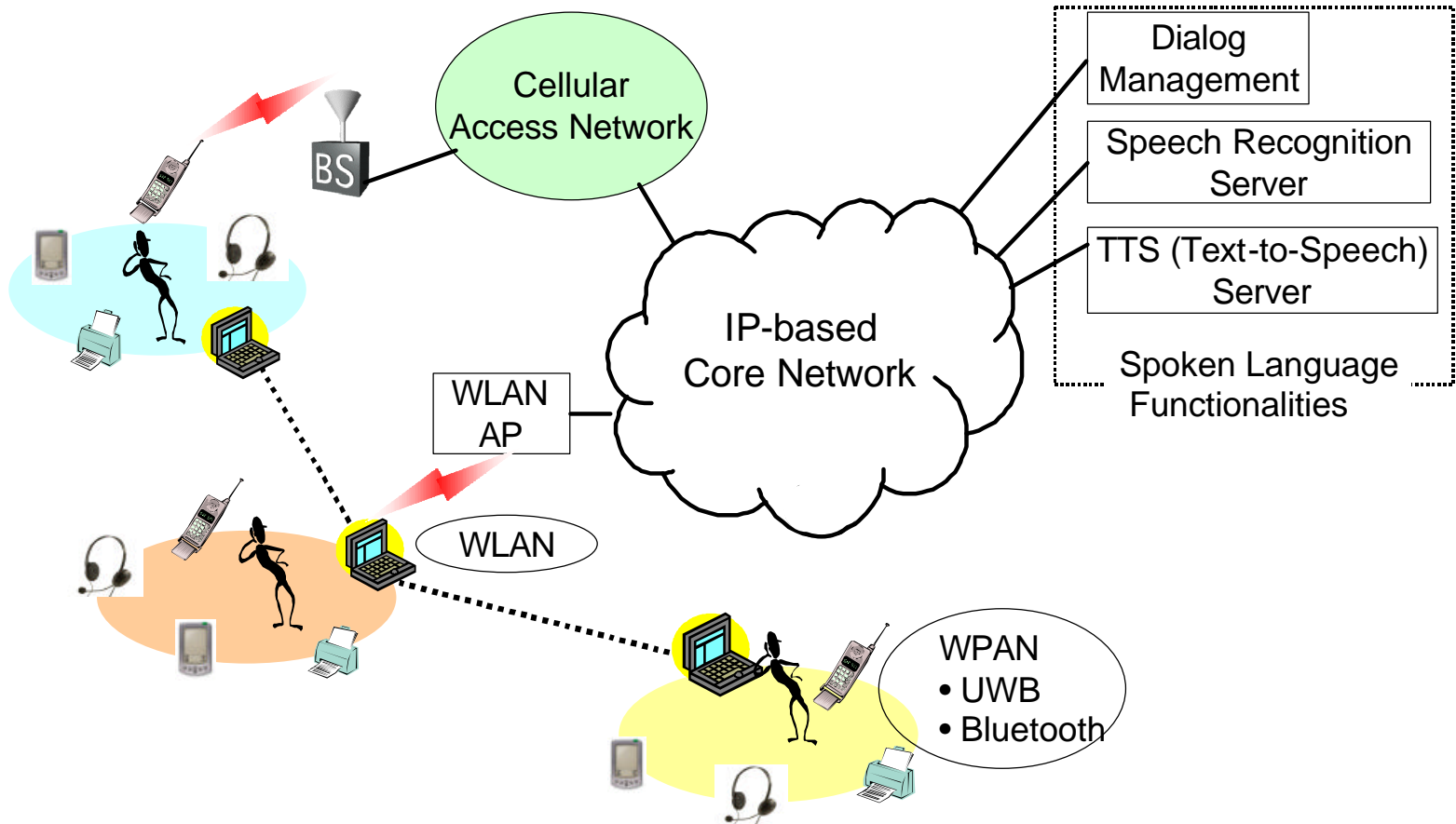
Kommentar

- Jo mere specialiseret ordforråd, jo dyrere.
- Systemerne er ret ens, med ca. 50.000++ ord +backup
- De genkender naturlig, flydende tale

Se også: www.voicerecognition.com

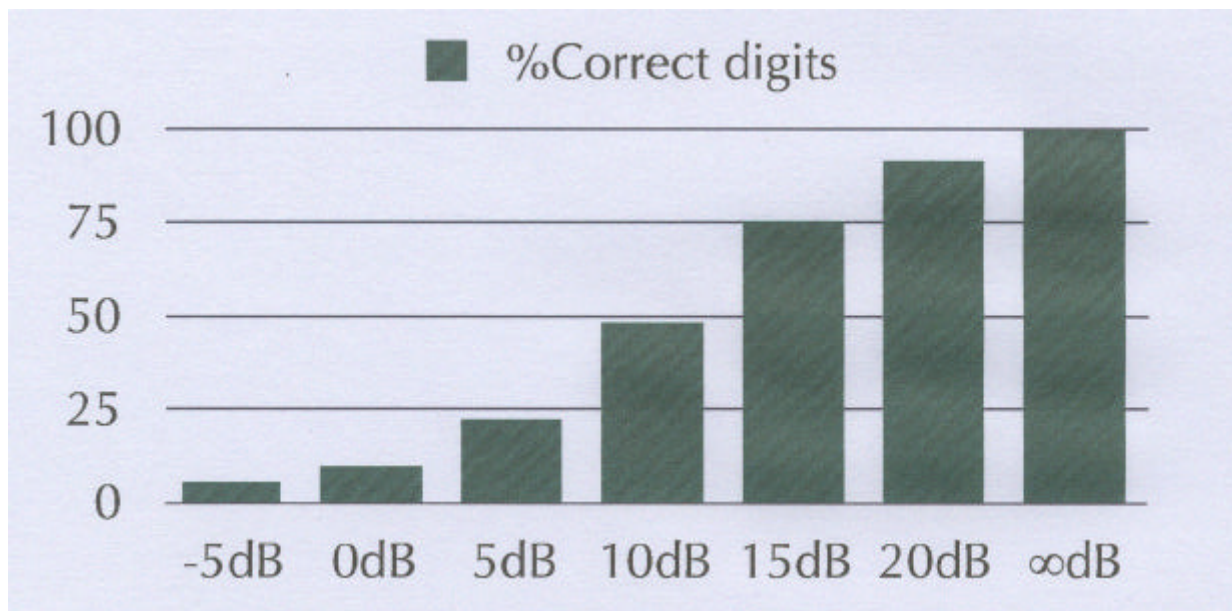


Distribueret processing (client-server)





DSR (Distributed Speech Processing):
Støj pga. Omgivelser og transmission (bit- og pakkefejl)





Projektets partnere er Aalborg Sygehus Patologisk Institut & IT-Sundhed, MaxManus samt Aalborg Universitet.

Problemer med **integrationen** af talegenkenderen.

	1. Samme dag		2. Dagen efter		3. Dagen efter*	
	Indhold	syntax	Indhold	syntax	Indhold	syntax
Indsatte fejl	68	66	79	59	70	67
Korrigerede fejl	36	46	29	25	13	34
%	53%	70%	37%	42%	19%	51%

Formål:

at undersøge, om der, ved at indføre talegenkendelse, kan opnås resource-besparelse i den daglige arbejdsrutine på Patologisk Institut på Aalborg Sygehus.



- Specifik dansk talegenkendelse (ikke generel) – f.eks. patologi
- Microsoft SQL Serverbaseret (MSDE)
- Recognition server
- Mange workstations kan tilkobles samme server
- Acoustic adaptation
- Context adaptation

- Demo ?

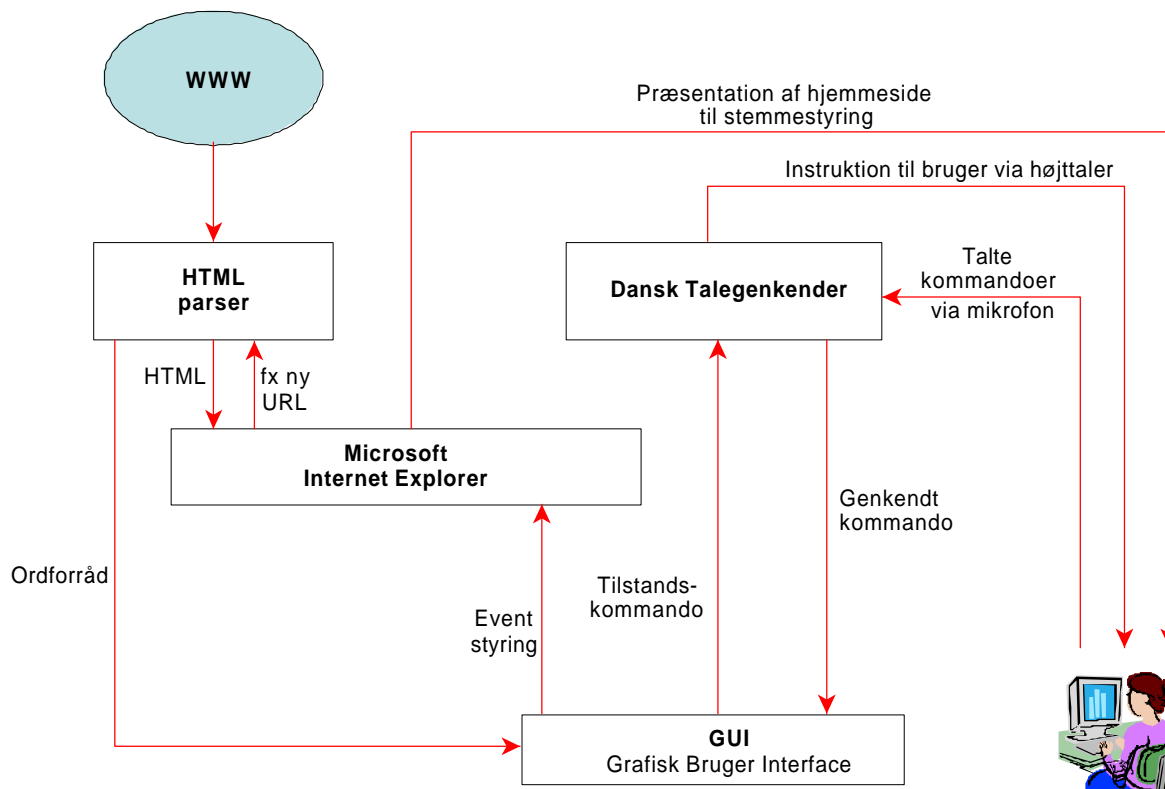


- Voice browsing for bevægelseshæmmede
- Begrænset navigation (browser kommandoer, HTML-link-aktivering, stemmestyret mus, bogstavering)
- På dansk
- Bevilget af MVTU til AAU, SPVV, Efaktum og HCH, Aalborg
- Efterfølger til adgangforalle.dk
- Tidsramme: Januar 2004 – Marts 2005
- Involverer databaseopsamling
- Talegenkender og database frit tilgængeligt efter projektets udløb
- Personer over 15 år
- Mindst to USB mikrofoner, headset-type



HTML parser

HTML sider fortolkes, tekst adskilles fra andre typer (fx billeder), identifikation af URL'er, samt udskillelse af hyperlink tekst, der forsynes med ekstra tags - velegnet til stemmestyring - og som benyttes til generering af grammatikker under genkendelsen.





- Teknologien er her i dag, og mange kommercielle systemer er allerede i brug
- Danske kommercielle produkter er ved (på vej til) at blive udviklet
 - kun få teknologiske barrierer
 - mangel på resurser=kapital
- Produkterne kommer i takt med at
 - programmerne forfines og standardiseres, træningsmaterialet øges og forbedres, anvendelserne opdages og udforskes
- Databaserne findes (akustiske og sproglige), men er bundet til omgivelserne (bil, GSM, fastnet)
- Fortsat behov for forskning mhp. øget robusthed (f.eks. støj, OOV), tilpasning (til netværk), tale som én ud af flere modaliteter, flersprogethed og "non-native" dansk