




- ### Agenda
- Dialog Systems
  - Data Collection
  - Recognition Understanding
  - Disfluency
  - Generation, Vocabulary
  - Dialog models
  - Spontaneous Data
  - Platforms
  - Evaluation
  - Error Handling
  - Challenges



## Classic systems

- Research systems
  - Voyager (1989)
  - ATIS (1992)
  - SUNDIAL (1993)
  - TRAINS (1996)
- Application
  - Philips Train Information (1995)
- Large Efforts
  - Communicator
  - Verbmobil








## The TRAINS Project: Natural Spoken Dialogue and Interactive Planning


Conversational Interaction and Spoken Dialogue Research Group

Project Leader  
James Allen  
TRAINS-91 was limited in scope but were important demonstration of the "doability" of spoken dialogue systems.


<http://www.cs.rochester.edu/research/cisd/projects/trains/>





## TRIPS


The Interpretation Manager (IM) interprets user input. It broadcasts the recognized speech acts and incrementally updates the Discourse Context.




The Generation Manager (GM) plans the specific content of utterances and display updates.

The Behavioral Agent (BA) plans system behavior based on its goals and obligations, the user's utterances and actions, and changes in the world state.


Figure 1: New Core Architecture  
James Allen et al "Towards conversational human-computer interaction," AI Magazine, 22(4), 2001






## Classic systems

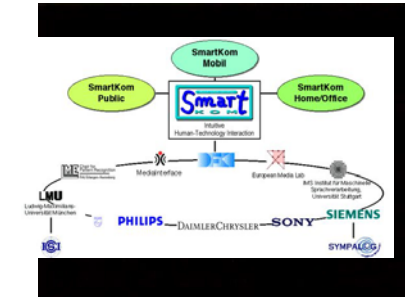

- Historic research systems
  - Voyager (1989)
  - ATIS (1992)
  - SUNDIAL (1993)
  - TRAINS (1996)
- Application
  - Philips Train Information (1995)
- Large Efforts
  - Communicator
  - Verbmobil
  - SmartKom






## SmartKom



Dialog-based Human-Technology Interaction by Coordinated Analysis and Generation of Multiple Modalities



## SmartKom


Dialog-based Human-Technology Interaction by Coordinated Analysis and Generation of Multiple Modalities

**Nordic Scene**

- Stockholm, Sweden
  - Waxholm (1993)
- Linköping, Sweden
  - LINLIN
- Göteborg, Sweden
  - TRINDI, GODIS
- Aalborg, Denmark
- Helsinki, Finland
- Trondheim, Norway

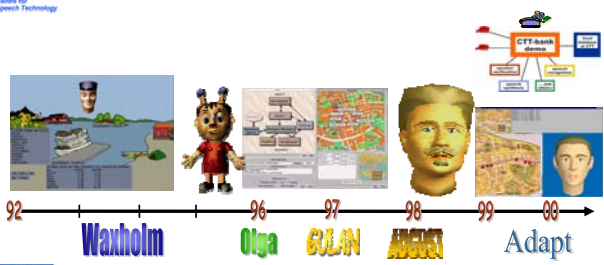
**The Waxholm interface**



**Some web pages on spoken dialogue systems**


- <http://www.cs.cmu.edu/~dbohus/SDS/index.html>
- <http://wwwhome.cs.utwente.nl/~schooten/vidiam/dialoguesystems/>
- <http://www.cs.cmu.edu/~dgroup/>
- <http://www.cs.cmu.edu/~dod/roundtable/>

**Dialog systems at KTH**



**Dialog systems at KTH**

2003



**Dialog systems at KTH**

2003



**Dialog systems at KTH**

2003

The HIGGINS domain

KTH logo

**AdApt multimodal dialog system**

Multimodal dialog system

Conversations about apartments for sale

Work together with a animated agent, Urban

Timeline: 92 Waxholm, 96 Olga, 97 GULAN, 98 AUGUST, 99, 00 Adapt

KTH logo

**Dialog Phenomena**

- "Har du inget billigare?"  
Implicit reference, ellipsis, context
- "Berätta mer om den andra lägenheten!"  
Meta-reference
- "Vad menar du med charmig?"  
Domain-question

KTH logo

**Simulation (Wizard-of-Oz)**

User

Human operator

KTH logo

**Wizard of Oz**

- How much does the wizard, WOZ, take care of
- The Complete System
- Parts of the system
  - Recognition
  - Synthesis
  - Dialog Handling
  - Knowledge Base
- Which demands on the WOZ
  - How to handle errors
  - Should you add information
  - What is allowed to say
- Which support does the WOZ have

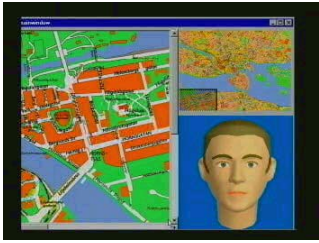
KTH logo

**Wizard-of-Oz data collection**

The Wizard's graphical interface

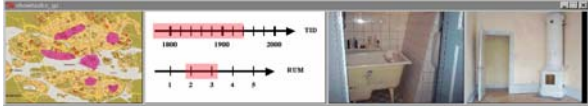
KTH logo

## Early demo




The screenshot shows a graphical user interface with a map of a city area on the left and a 3D head model on the right. The map displays various buildings and streets in a stylized, colorful manner. The head model is a simple, realistic-looking face.

## Pictorial scenarios



The screenshot displays three different visual elements: a map of a city area on the left, a timeline or calendar view in the center, and a photograph of a room interior on the right. The timeline shows dates from 1990 to 2000, and the room interior shows a simple living space with a desk and a chair.

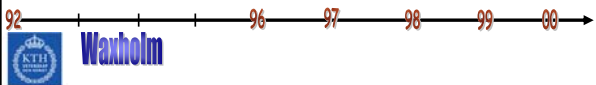
## Adapt - demonstration of "complete" system



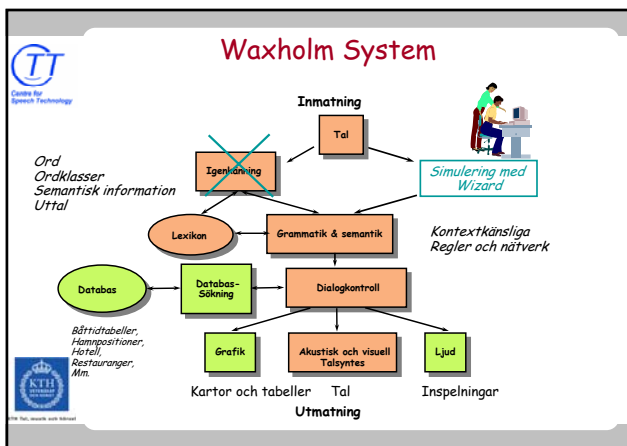
The screenshot shows a graphical user interface with a map of a city area on the left and a 3D head model on the right. The map displays various buildings and streets in a stylized, colorful manner. The head model is a simple, realistic-looking face.

## The Waxholm Project

- tourist information
  - Stockholm archipelago
  - time-tables, hotels, hostels, camping and dining possibilities.
- mixed initiative dialogue
  - speech recognition
  - multimodal synthesis
- graphic information
  - pictures, maps, charts and time-tables.




The timeline shows the years 1992, 1996, 1997, 1998, 1999, and 2000. The word "Waxholm" is written in a stylized font below the timeline.

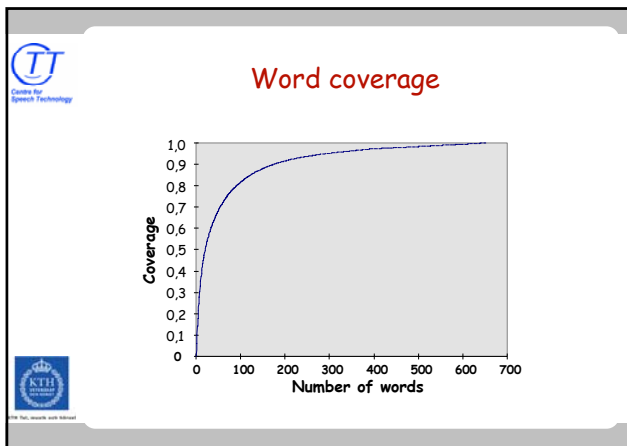


## Waxholm Database

- About 70 subjects (9200 words)
- Phonetically transcribed
- Examples from the Waxholm system
  - Five different speakers
    - EJ KR GO LN MK

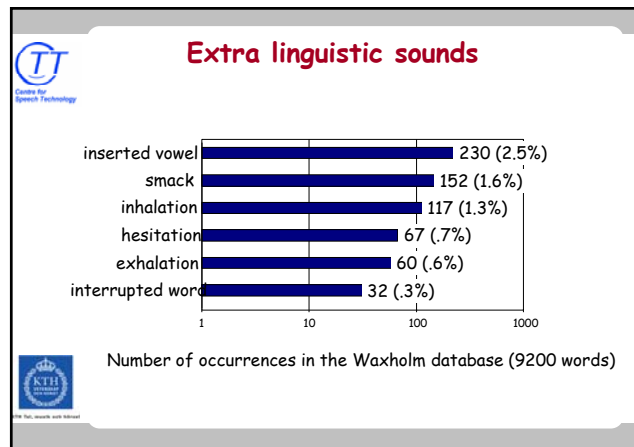
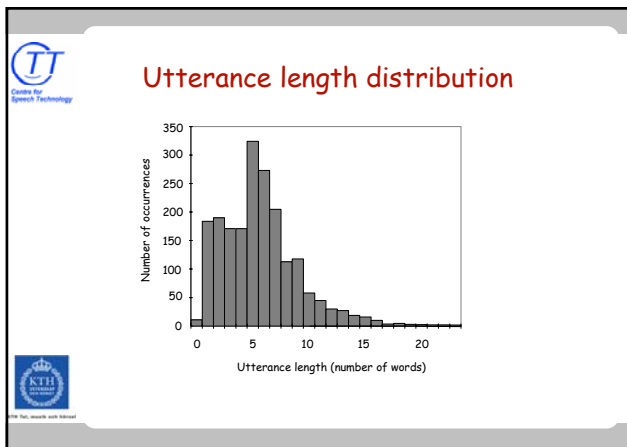


The icons represent the five different speakers: EJ, KR, GO, LN, and MK. Each speaker is represented by a small icon of a person's head.

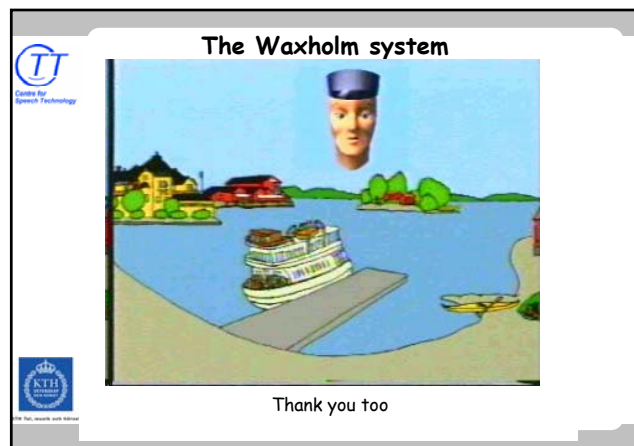


**Lexicon - transcription**

Word	freq.	Word	freq.
skärgården	69	skulle	23
SJA3RGÅ:2N	15	SKK"ULE0	26
SJA3RGÅ:2N	6	SKK"U	3
SJA3RGÅ:2N	5	SKK"UL	2
SJA3RGÅ:2DdE0N	5	SKLLE0	1
SJA3RGÅ:2DdE0N	4	SKK"UE0	1
SJA3RGÅ:2DE0N	4		
SJA3RGÅ:2DE0N	4		
SJA3RGÅ:2DE02N	3		
2S'A3RGÅ:2N	3		
SJA3RGÅ:N	2		
SJA3RGÅ:2DdE02N	2		
SJA3RGÅ:2DdE02N	2		
SJA3RGÅ:2DdE0N	2		
SJA3RGÅ:2DdE0N	1		
SJA3RGÅ:N	1		
SJA3RGÅ:2DdE02N	1		
SJA3RGÅ:2DE0N	1		
SJA3RGÅ:2DE02N	1		
SJA3RGÅ:2DE0Nv	1		
SJA3RGÅ:2DE0N	1		
SJA3RGÅ:2DE0N	1		
SJA3RGÅ:2DdE0N	1		

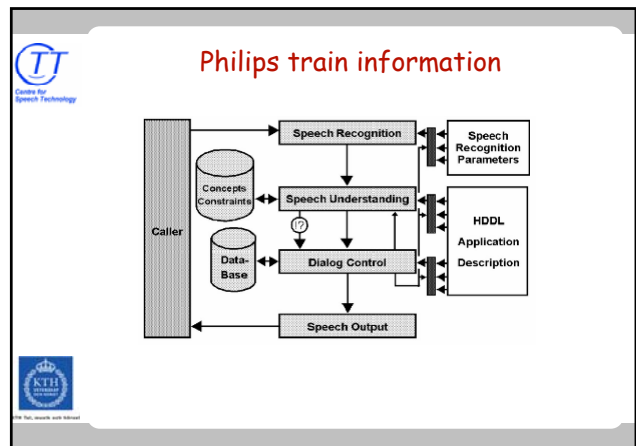


Three years later....

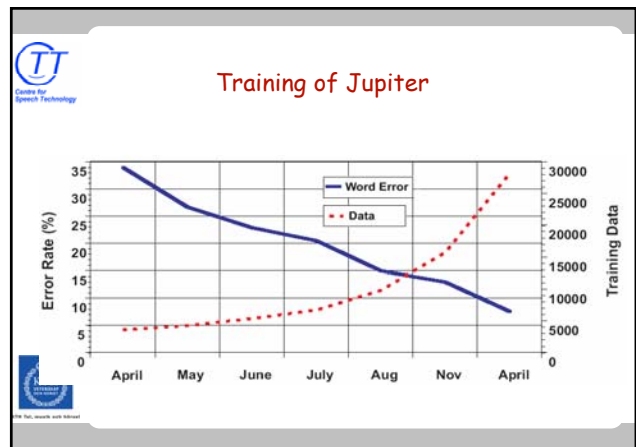


**Instead of WOZ  
"Bootstrap" the system**

- Make a **simple but complete** system and evaluate
- Spread the information....
- Collect data
- Upgrade the system



**WebGALAXY Display**



Even generic databases are important

[Swedia <http://swedia.ling.umu.se/>  
SpeechDat](http://swedia.ling.umu.se/SpeechDat)

**SpeechDat**

**Swedish dialects**

"Flyget, tåget och bilbranschen tävlar om lönsamhet och folkets gunst".

Född i  
USA  
ex-Jugoslavien



**Speech understanding some aspects**

- Bigram
- Tight coupling
- Keyword spotting
- Phrase spotting
- Full grammatical and semantic analysis
- OOV out of vocabulary

**Knowledge sources - Evaluation**

Acoustic analysis  
Syntactic analysis  
Semantic analysis  
Dialog state  
Dialog Context

Confidence  
Expectation  
Filter

**Multi-level analysis**

SCORE =  $S_{pnode} + S_{pterminal} + S_{pword} + f(\text{length})$

**I want to go.....**

Parser score

12.26	Jag vill åka från Stockholm till Vaxholm. <i>I want to go from Stockholm to Vaxholm.</i>
11.99	Jag vill åka till Vaxholm från Stockholm. <i>I want to go to Vaxholm from Stockholm.</i>
10.01	Jag vill åka till Vaxholm. <i>I want to go to Vaxholm.</i>
9.85	Jag skulle vilja åka till Vaxholm. <i>I would like to go to Vaxholm.</i>
5.30	Jag vill åka. <i>I want to go.</i>
3.17	När går det en båt till Vaxholm? <i>When does a boat go to Vaxholm?</i>
-1.32	När går båten till Vaxholm? <i>When does the boat go to Vaxholm?</i>
-1.95	Jag vill åka till mamma. <i>I want to go to my mother.</i>

**Robust Analysis**

match

CA TYPE: ASSESS

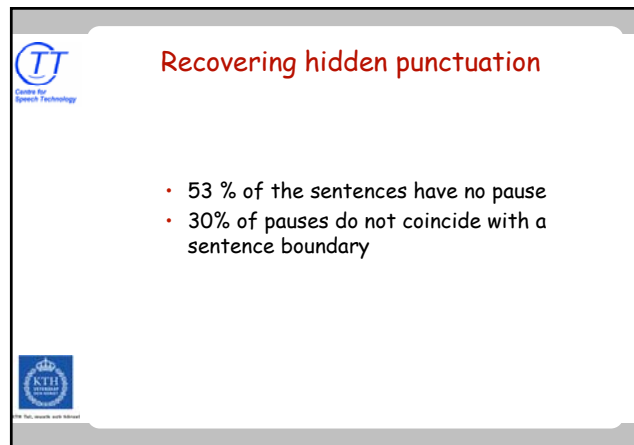
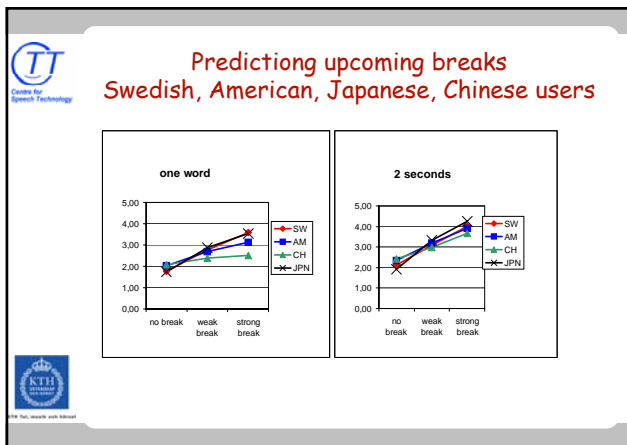
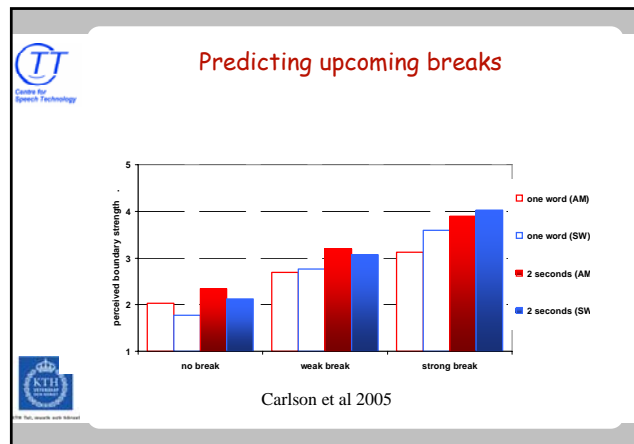
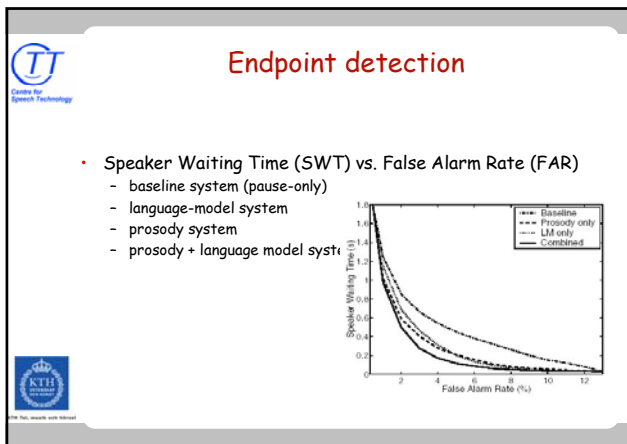
- Robust interpretation
  - Using grammar to automatically detect non-expected words between and inside phrases
  - Performs better than keyword-spotting for detecting erroneous content-words
  - Skantze, G. & Edlund, J. (2004). *Robust interpretation in the Higgins spoken dialogue system.*

**Spontaneous Speech: How People Really Talk and Why Engineers Should Care**

- Recovering hidden punctuation
- Coping with disfluencies
- Allowing for realistic turn-taking
- Hearing more than words

Elizabeth Shriberg (Interspeech 2005)

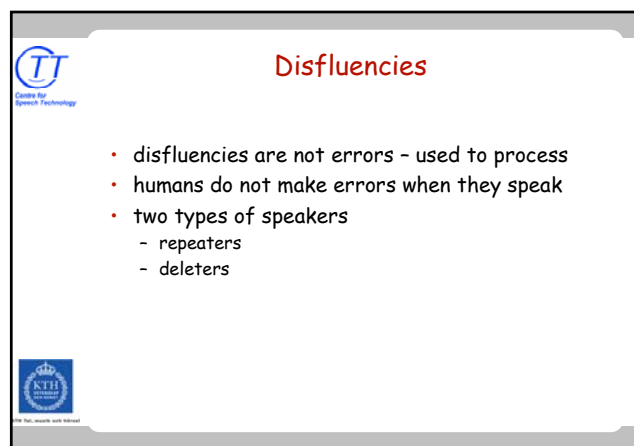




**Disfluencies**

Type	Example
Filled pause	jag <u>hum</u> tycker om glass
Repetition	<u>jag</u> jag tycker om glass
Insertion	jag tycker <u>om</u> inte om glass
Restart	<u>kan du</u> jag tycker om glass
Substitution	<u>vi</u> tycker jag tycker om glass

Carlson et al 2005



S. Oviatt

## 'Disfluency rate'

- **Human - Human**
  - Two person telephone
  - Two person direct
  - One person
- **Human - Machine**
  - Computer interaction

High

↑

Low

## Distribution of Disfluencies

Switch board data, Liz Shriberg, Thesis, SRI

## Disfluency examples from Adapt

rättelse	det är lite för ... lite för sent <b>tidigt</b> finns det nån <b>ehm</b> ... likande lgh ... in/~ området med som är byggd på 1800talet
avbrutet	hur se/ <b>eh</b> ... är kaklet <b>eh</b> ... utrustat
pauser	uhm ... högt frill fok och <b>eh</b> ... kanske någon kakelugn ... och balkong gärna i i söderläge
feluttal	är den <b>eh</b> nyreda~ nyrenoverad
förlängning	<b>hauuun</b> ser gatan ut
rättelse/term	jag vill gärna ha en lägenhet med ... utstikt ... nej med balkong

## Disfluencies in half of the Adapt corpus

22% of all utterances disfluent  
6% of all words disfluent

Percentage disfluent words

Utterance lengths

Percentage disfluent words in turns with five to nine words

Individual users

## Utterance Generation

- Predefined utterances
- Frames with slots
- Generation based on grammar and underlying semantics

## System Utterances

- The output should reflect the system's vocabulary and linguistic capability
  - the users adapt
- Short utterances
  - The users adapt
- Good error messages
  - Use words and phrases the system can handle

**User answers to questions?**

The answers to the question:  
**"What weekday do you want to go?"**  
 (Vilken veckodag vill du åka?)

- 22% **Friday** (fredag)
- 11% **I want to go on Friday** (jag vill åka på fredag)
- 11% **I want to go today** (jag vill åka idag)
- 7% **on Friday** (på fredag)
- 6% **I want to go a Friday** (jag vill åka en fredag)
- - **are there any hotels in Vaxholm?**  
 (finns det några hotell i Vaxholm)

**User answers to questions?**

The answers to the question:  
**"What weekday do you want to go?"**  
 (Vilken veckodag vill du åka?)

- 22% **Friday** (fredag)
- 11% **I want to go on Friday** (jag vill åka på fredag)
- 11% **I want to go today** (jag vill åka idag)
- 7% **on Friday** (på fredag)
- 6% **I want to go a Friday** (jag vill åka en fredag)
- - **are there any hotels in Vaxholm?**  
 (finns det några hotell i Vaxholm)

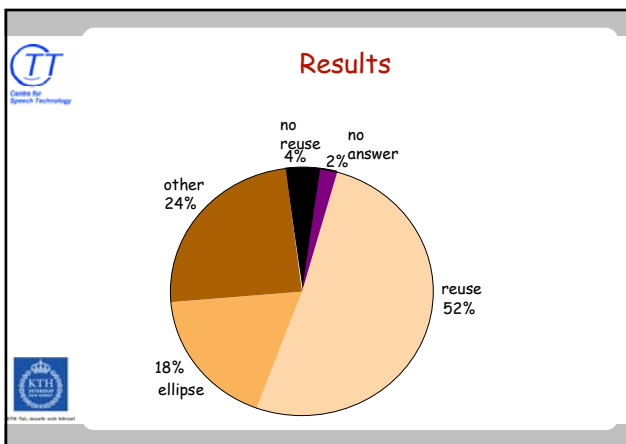
**Pairs of alternative main verbs**

höra- lyssna	(listen - hear)
vandra- ströva	(hike-stroll)
köpa - handla	(shop-buy)
se på - gå på	(watch- go to)
föredrar -tycker mest om	(prefer -like the most)
testa - pröva	(test-try)

**Exemple of questions and answers**


Hur ofta **åker** du utomlands på semester? Hur ofta **reser** du utomlands på semester?

jag åker en gång om året kanske	jag reser en gång om året utomlands
jag åker ganska sällan utomlands på semester	jag reser inte ofta utomlands på semester det blir mera i arbetet
jag åker nästan alltid utomlands under min semester	jag reser reser utomlands på semester varannat år
jag åker ungefär 2 gånger per år utomlands på semester	jag reser utomlands en gång per semester
jag åker utomlands nästan varje år	jag reser utomlands på semester ungefär en gång per år
jag åker utomlands på semester varje år	jag brukar resa utomlands på semester åtminstone en gång i året
jag åker utomlands ungefär en gång om året	en gång per år kanske
jag är nästan aldrig utomlands	en gång vart annat år
en eller två gånger om året	varje år
en gång per semester	vart tredje år ungefär
kanske en gång per år	nu för tiden inte så ofta
ungefär en gång per år	varje år brukar jag åka utomlands
åtminstone en gång om året	
nästan aldrig	




**Lessons**

- subjects adapt their lexical choices to system questions
- less than 5% of the cases an alternative main verb is used in the answer
- adaptive language model and lexicon in the recognizer




Centre for Speech Technology

## Dialog Model




KTH, research with passion




Centre for Speech Technology

## Human- machine interaction

- **Initiative**
  - system/user
- **Who is the user**
  - First time?
- **Terminology**
  - joint vocabulary
- **Do you accept barge in?**
  - Has the user understood what was said?
- **Can the user teach the system?**




KTH, research with passion




Centre for Speech Technology

## Modalities

- **Who are you talking to**
  - system
  - Animated character
- **How is the information presented**
  - Text, tables, pictures
  - Synthetic speech
- **Can you both talk and point**




KTH, research with passion




Centre for Speech Technology

## Spoken dialog system

- **Finite-state based systems**
  - dialog and states explicitly specified
- **Frame based systems**
  - dialog separated from information states
- **Agent based systems**
  - model of intentions, goals, beliefs




KTH, research with passion



Centre for Speech Technology

## Dialog model

- **Domain dependent model**
  - Rules, networks, stack
- **Separate models for the dialog turns and the semantics**
  - For example Question/answer
- **Reference Handling**



KTH, research with passion



Centre for Speech Technology

## voiceXML

voiceXML FORUM Sponsors and Promoters

- **Sponsors - 4**
  - IBM, AT&T, Lucent and Motorola
- **Promoters - 23**
  - @VoiceGenie Technologies Inc.
  - Alcatel
  - AnyDevice.com, Inc.
  - Brence, Inc.
  - Cisco Systems
  - Converse Network Systems
  - Enuncia Communications
  - Hitachi, Ltd. Central Research Lab
  - Huawei Technologies Company Ltd
  - Lernout & Hauspie
  - Milo
  - MobileWebSurf
  - Mockingbird Networks
  - Nhnancements Technologies, Inc.
  - Nuance Communications
  - Oki Electric Industry Co., Ltd
  - Oracle
  - SpeechHost, Inc.
  - SpeechWorks International
  - Telera
  - VercomNet BV
  - Vocollect, Inc.
  - Voxeo

<http://www.voicexml.org/>



KTH, research with passion

**Nuance Voyager**

**SpeechObjects™ & VoiceXML**

Listen for These Nuance Voyager Advantages

- Profiles** - Contains personalized information that can be shared with different sites.
- Bookmarks** - A personal list of frequently used voice sites or phone numbers.
- Hotword** - The word a user says at any time to get back to Voyager. He can then visit another voice site or use other Voyager features.
- Search** - A powerful feature that provides the option to ask for a specific name or business or search the yellow pages by category.
- Browsing** - The ability to go to voice enabled Internet content sites.
- Navigation** - The ability to move back and forward between voice sites that have already been visited during a session.
- Voicelinks** - Voicelinks operate the same way as hyperlinks on the Internet, whereby a phrase enclosed in the voicelink sound indicates that a user can say the phrase and be link to another Voice Site.
- Verification** - Uses a sophisticated voice printing technology to secure e-commerce transactions. Voice printing with knowledge verification is extremely secure. [Click here](#) for information about Nuance Verifier.

**Waxholm Topics**

**TIME\_TABLE** Task: get a time-table.  
Example: När går båten? (When does the boat leave?)

**SHOW\_MAP** Task: get a chart or a map displayed.  
Example: Var ligger Vaxholm? (Where is Vaxholm located?)

**EXIST** Task: display lodging and dining possibilities.  
Example: Var finns det vandrarhem? (Where are there hostels?)

**OUT\_OF\_DOMAIN** Task: the subject is out of the domain.  
Example: Kan jag boka rum. (Can I book a room?)

**NO\_UNDERSTANDING** Task: no understanding of user intentions.  
Example: Jag heter Olle. (My name is Olle)

**END\_SCENARIO** Task: end a dialog.  
Example: Tack. (Thank you.)

**Dialogue control - state prediction**

Dialog grammar specified by a number of **states**  
Each state associated with an **action**  
database search, system question... ..

Probable state determined from **semantic features**  
Transition **probability** from one state to state  
Dialog control **design tool** with a graphic interface

**Semantic Frame**

Current functions: /TO-PLACE Q-VERBAL SUBJECT FROM-TIME/  
Current meaning: /MOVE BOAT PORT QUANT/

History functions: /TO-PLACE Q-VERBAL SUBJECT FROM-TIME/  
History meaning: /MOVE BOAT PORT QUANT/

(FROM-TIME AFTER\_TIME "04"/)  
(FROM-TIME BEFORE\_TIME "06"/)  
(SUBJECT "båten"/BOAT/)  
(Q-VERBAL "går"/MOVE/)  
(TO-PLACE "vaxholm"/PORT/)

proposed topic TIME\_TABLE


**Topic selection**

FEATURES	TIME TABLE	SHOW MAP	FACILITY	NO UNDERSTANDING	OUT OF DOMAIN	END
OBJECT	.062	.312	.073	.091	.067	.091
QUEST-WHEN	.188	.031	.024	.091	.067	.091
QUEST-WHERE	.062	.688	.390	.091	.067	.091
FROM-PLACE	.250	.031	.024	.091	.067	.091
AT-PLACE	.062	.219	.293	.091	.067	.091
TIME PLACE	.312	.031	.024	.091	.067	.091
OOD	.062	.200	.500	.091	.067	.091
END	.062	.031	.122	.091	.933	.091
HOTEL	.062	.031	.024	.091	.067	.909
HOSTEL	.062	.031	.488	.091	.067	.091
ISLAND	.062	.031	.122	.091	.067	.091
PORT	.333	.556	.062	.091	.067	.091
MOVE	.125	.750	.244	.091	.067	.091
	.875	.031	.098	.091	.067	.091


$$\underset{i}{\operatorname{argmax}} \{ p(t_i | F) \}$$

**Topic prediction results**


Condition	All (%)	"no understanding" excluded (%)
complete parse	3.1	2.9
raw data	12.9	8.8
no extra linguistic sounds	12.7	8.5




## How may I help you?



- Callers are routed to support staff using Natural Voices technology, AT&T Consumer Services' How May I Help You? (HMIHY).
- The HMIHY system was deployed in 2001, and by the end of the year, it was handling more than 2 million calls per month.
- Allen Gorin et al.
- > Demos on web page.....
- [C:\My Web Sites\howmayihelptou\www.research.att.com/~jwright/hmihy/samples.html](http://www.research.att.com/~jwright/hmihy/samples.html)




KTH, research with passion




## Interaction control

- Conversation
  - Exchange of information
  - Control of the exchange of information
- Turn-taking
  - Control of the 'floor'
- Feedback
  - Perception, attention, understanding, attitude...
- Dialogue systems needs both turn-taking and feedback




KTH, research with passion




## Human-human conversations

Act	Customer		Agent	
	Freq.	Words	Freq.	Words
Acknowledge	47.9	2.3	30.8	3.1
Request	29.5	9.0	15.0	12.3
Confirm	13.1	5.3	11.3	6.4
Inform	5.9	7.9	27.8	12.7
Statement	3.4	6.9	15.0	6.7




KTH, research with passion

Statistics of turns in a movie domain (from Flammia).




## Conversational "grunts"

- Grunts occur an average of once every 5 seconds in American English conversation. (Nigel Ward, 2000)
- In Switchboard database
  - um was the 6th most frequent item (after I, and, the, you, and a), (Nigel Ward, 2000)
  - the four items uh, uh-huh and um and um-hum accounted for 4% of the total (Picone et al. 1998).



KTH, research with passion



## Notation

abbreviation and function/position

back back-channel

fill filler, including various things that occur utterance- or turn- initially


dis disfluency marker

is isolate, produced when neither person has the turn, typically more self-directed than other-directed


rs response to direct question or high-rise statement

c confirmation, in response to a back-channel

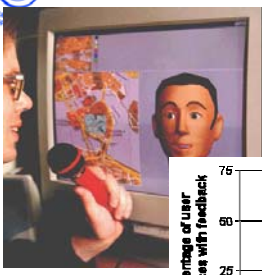
o other, including clause-final items, items that occur in quotations, and items whose function is obscure



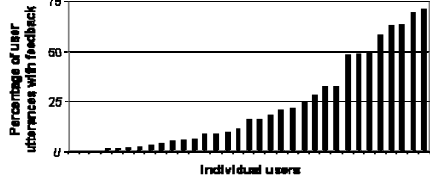
KTH, research with passion




## User studies



- Turn-taking
- Interaction
- Positive and Negative User Feedback
- User reactions





KTH, research with passion

### Positive and negative feedback

System	User	Feedback
This house was built in 1600	Oh!	NegativeAttitude
This building was constructed in 1861	Yes yes that's right.... is there a tiled stove there too	Positive Attention
This apartment has a fireplace	Yes that's all right too ..... how high is the building	Positive Attitude
This apartment is on the first floor	Okay .... and I see it is close to the German church there	Positive Attention
I don't know anything about such things	Well okay..... yes but I think I'm happy with that	NegativeAttitude

94% of the subjects used feedback at least once  
 65% of the feedback turns were labeled as positive  
 18% of all user utterances contained feedback  
 6% of the feedback occurred in a separate turn

### Parameter settings to create different stimuli

	Affirmative setting	Negative setting
Smile	Head smiles	Head has neutral expression
Head movement	Head nods	Head leans back
Eyebrows	Eyebrows rise	Eyebrows frown
Eye closure	Eyes close a bit	Eyes open widely
F0 contour	Declarative intonation	Interrogative intonation
Delay	Immediate reply	Slow reply

### The August system

- Stockholm (events and general information)
- Yellow pages
- KTH and speech technology
- August Strindberg
- Greetings and social utterances
- Comments about the system capabilities and the discourse

### Shallow semantic analysis

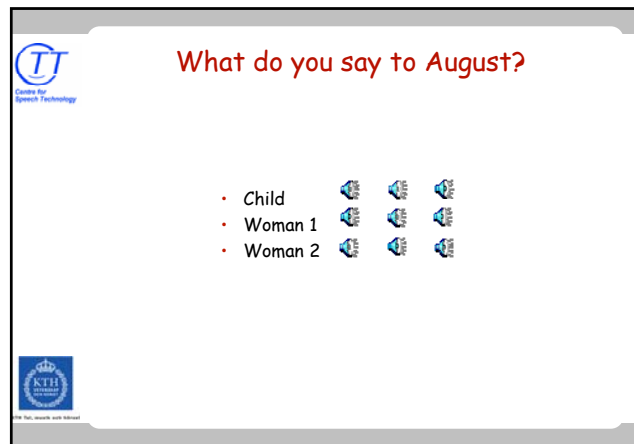
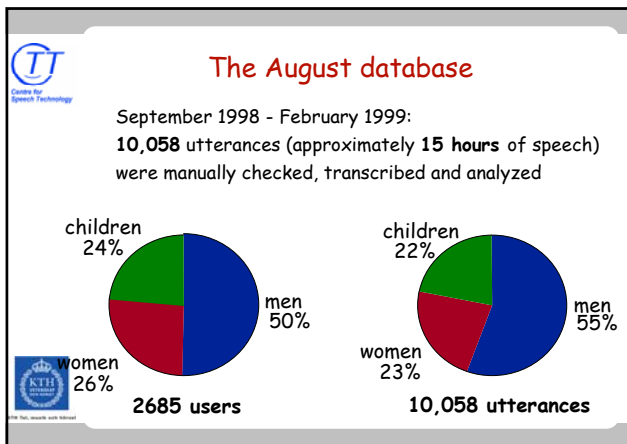
- Input
  - word sequences
  - semantic features from lexicon
- Output
  - Acceptable utterance? yes/no
  - Predicted domain
    - strindberg, stockholm, yellow pages....
  - Feature:value representation
    - object:restaurant, place:mariafortet
- Trained on tagged N-best lists and lexicon

### The set-up in Kulturhuset

### A sample video of the system environment

My life cannot be measured in terms of days and years!





**Utterance types in the August database**

Socializing	Info-seeking
Social	Domain
Insult	Meta
Test	Facts

**Socializing categories**

<b>Social</b>	<i>Hello August!</i> <i>That's a nice moustache!</i> <i>Would you like to go out with me tonight?</i>
<b>Insult</b>	<i>You are stupid!</i> <i>Is your brain too small?</i> <i>You have a sausage brain!</i>
<b>Test</b>	<i>What is my name?</i> <i>I want to rent a refrigerator</i> <i>What is the colour of your hair?</i>

**The info-seeking categories**

<b>Domain</b>	<i>How many books did Strindberg write?</i> <i>What can you study at KTH?</i> <i>Where are the restaurants on Kungsgatan?</i>
<b>Meta</b>	<i>What can I ask you?</i> <i>August answer my question I know you know everything</i> <i>Then I will speak at the same time as I hold down the button - what is your name, agent</i>
<b>Facts</b>	<i>What's the capital of Finland?</i> <i>What is two times two?</i> <i>How many people live in Madrid?</i>

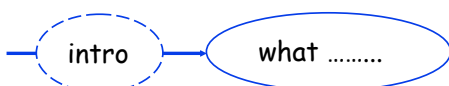
**User utterance categories during the first six dialogue turns**

category	children	women	men
only socializing	34%	20%	20%
only info-seeking	28%	39%	34%
from socializing to info-seeking	31%	35%	43%
alternating	7%	6%	3%

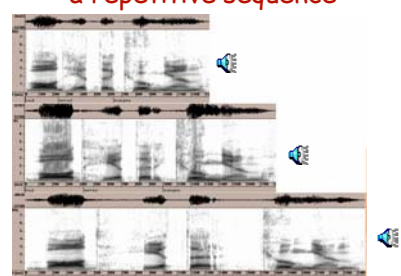
The statistics are based on the first utterances (up to six) from all users that said more than two utterances to the system

**What ..... ?**

- 334 utterances include "what"
  - only 75 have "what" in initial position
- 99 "what is your name"
  - all in final utterance position
  - only 13 initiate an utterance

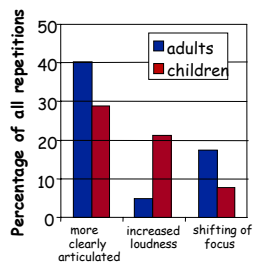


**An example of a repetitive sequence**



The utterance "Vad heter kungen?" (What is the name of the king?) as original input (top) and repeated twice by the same user

**Features in repetition**



Feature	adults (%)	children (%)
more clearly articulated	40	30
increased loudness	5	22
shifting of focus	18	8

**Some lessons for recognition**

- lexical entrainment
  - use both user input and system output
- adaptive to
  - application
  - user
  - dialog
- use three recognition systems in parallel
  - continuous speech (default)
  - word by word (error resolution)
  - continuous syllables (confidence)

**The August system**

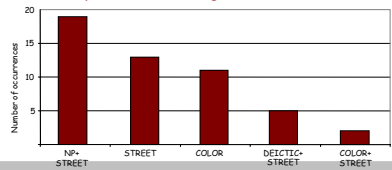


Strindberg was married three times!

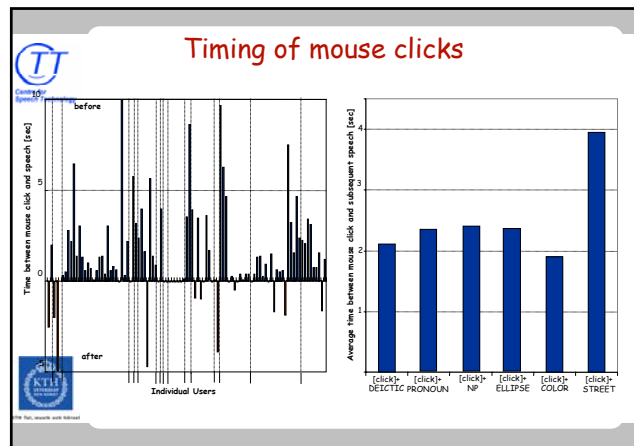
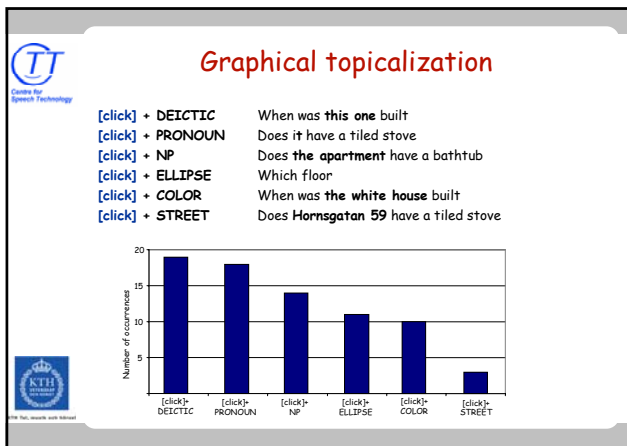
**Verbal topicalization**

About 10% of all requests for information about a specific apartment contained a topicalized reference, 4% contained a verbal topical reference and 6% a preceding graphical reference

Reference Type	Example
NP+STREET	The apartment on Sankt Eriksgatan - which floor is it on
STREET	Österlånggatan 24 - does it have a balcony
COLOR	The green apartment - does it have a balcony
DEICTIC+STREET	This one-on Heleneborgsgatan - does it have a bathtub
COLOR+STREET	The yellow one on Kocksgatan 20 - does it have a balcony



Reference Type	Number of occurrences
NP+STREET	19
STREET	13
COLOR	11
DEICTIC+STREET	5
COLOR+STREET	2

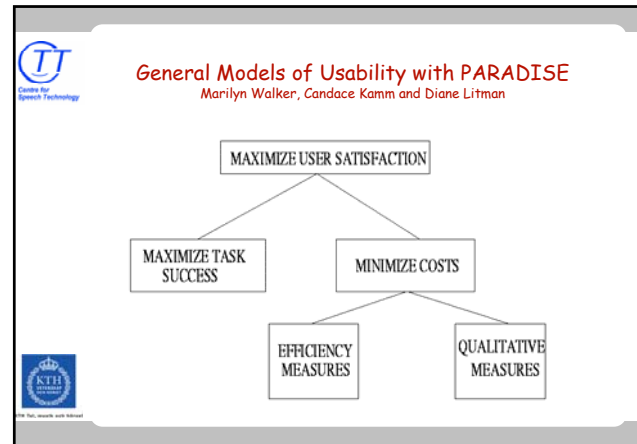
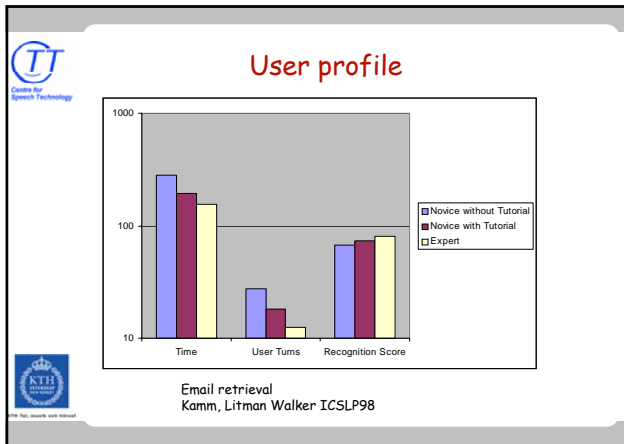


- ### Evaluation
- Phonetic analysis
  - Word understanding Synthesis/Recognition
  - Domain Dependent
    - Vocabulary Syntax
  - System Feedback
  - "Task completion"
  - How long time
  - How many turns
  - Happy and satisfied users

- ### Evaluation efforts
- Some projects
    - NIST
    - SAM
    - COCOSDA
    - EAGLES
    - DISC

- ### NIST-evaluations
- NIST - National Institute of Standards and Technology (USA)
    - <http://www.nist.gov/speech/>
  - Areas
    - Communicator (Intelligent Conversational Interfaces, 2000-)
    - Speech Recognition (English, Spanish, Mandarin)
      - broadcast news (1996-1999)
      - conversational telephone speech (1997-)
    - Topic Detection and Tracking (1998-, English, Mandarin)
    - Information Extraction - Entity Recognition (1999-)
    - Spoken Document Retrieval (1997-2000)
    - Speaker Recognition (1996-)

- ### DISC
- Spoken Language Dialogue Systems and Components:  
Best practice in development and evaluation
  - Partners and people
    - Natural Interactive Systems Laboratory (NIS), Denmark
    - Centre National de la Recherche Scientifique (CNRS-LIMSI) France
    - Universität Stuttgart, Germany
    - Kungliga Tekniska Högskolan (KTH), Sweden
    - Vocalis Ltd, England
    - Daimler-Chrysler AG, Germany
    - ELSNET, Europe
  - URL: [www.disc.dk](http://www.disc.dk)




- ### Paradise User Satisfaction
- I found the system easy to understand in this conversation. (TTS Performance)
  - In this conversation, I knew what I could say or do at each point of the dialogue. (User Expertise)
  - The system worked the way I expected it to in this conversation. (Expected Behaviour)
  - Based on my experience in this conversation using this system to get travel information, I would like to use this system regularly. (Future Use)

- ### Evaluation metrics
- Dialog Efficiency Metrics: Total elapsed time, Time on task, System turns, User turns, Turns on task, time per turn for each system module



- ### Evaluation metrics
- Dialog Efficiency Metrics: Total elapsed time, Time on task, System turns, User turns, Turns on task, time per turn for each system module
  - Dialog Quality Metrics: Word Accuracy, Sentence Accuracy, Mean Response latency, Response latency variance

- ### Evaluation metrics
- Dialog Efficiency Metrics: Total elapsed time, Time on task, System turns, User turns, Turns on task, time per turn for each system module
  - Dialog Quality Metrics: Word Accuracy, Sentence Accuracy, Mean Response latency, Response latency variance
  - Task Success Metrics: Perceived task completion, Exact Scenario Completion, Any Scenario Completion



## Evaluation metrics



- Dialog Efficiency Metrics: Total elapsed time, Time on task, System turns, User turns, Turns on task, time per turn for each system module
- Dialog Quality Metrics: Word Accuracy, Sentence Accuracy, Mean Response latency, Response latency variance
- Task Success Metrics: Perceived task completion, Exact Scenario Completion, Any Scenario Completion
- User Satisfaction: Sum of TTS performance, Task ease, User expertise, Expected behaviour, Future use.

## Prediction of satisfaction?

$$\text{PERFORMANCE} = .25 \text{ MRS} + .33 \text{ COMP} - .33 \text{ HELP}$$



MRS = mean recognition score  
 COMP = perceived completion  
 HELP = number of help messages  
 PERFORMANCE = User satisfaction  
 Covers 41.3% of the variance

## Dialog Management in MIMIC



- Initiative modeling
  - distribution of system initiatives
- Goal selection
  - goal that the system attempts to reach
- Strategy selection
  - dialog acts depending on initiative distribution

MIMIC: An Adaptive Mixed Initiative Spoken Dialogue System for Information Queries Jennifer Chu-Carroll, NAACL 2000



## Initiative - Cue detection

- Discourse cues
  - TakeOverTask
    - when user gives more info than needed
  - NoNewInformation
    - no progress towards task completion

## Adaptation of the dialog


- Evaluate the dialog continuously
  - Do the system and the user have the same goal
  - Who takes the initiative
- Error handling
  - Analysis and repair





## Robust to the understanding errors

- Breakdowns of communication seen as inherent properties of the activity
- Ability to keep the dialogue going


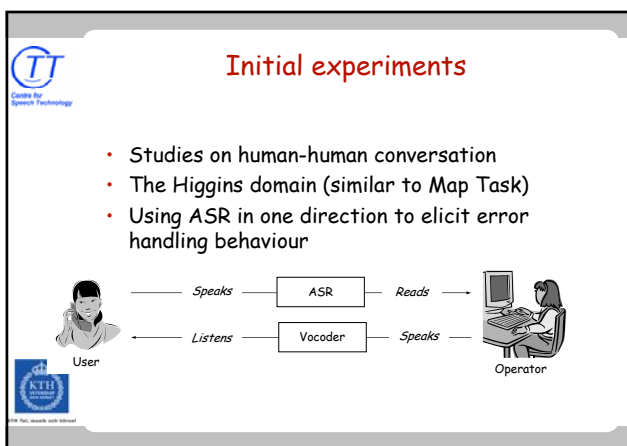
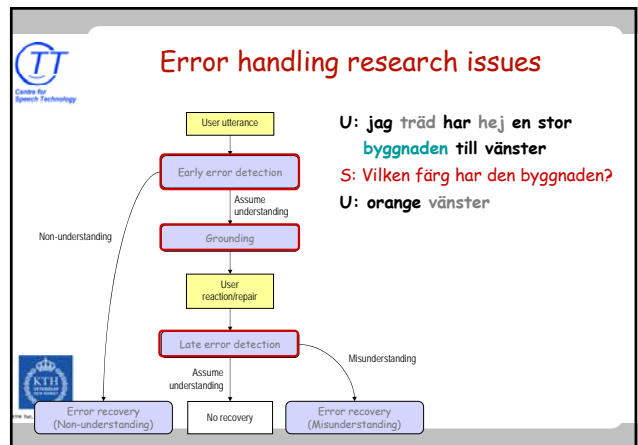

(Martinovski&Traum 2003)




## Errors in spoken dialogue systems

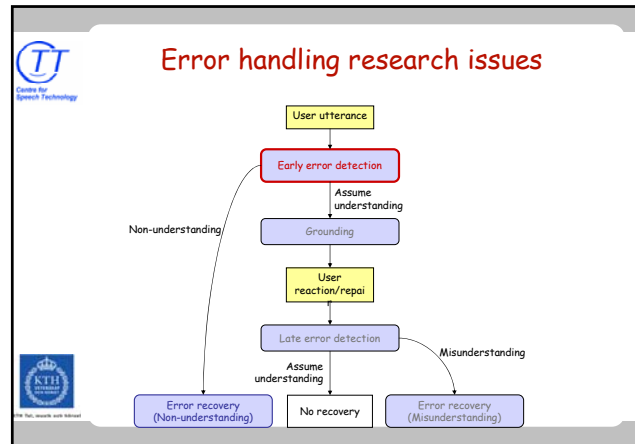
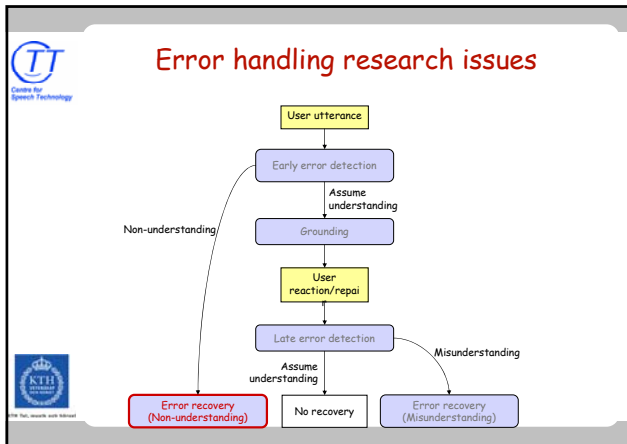
- What is an error?
  - A deviation from an expected output from a system, module or process
- Deviation from what?
  - What is written in the requirement specification
  - What a human "wizard" would do
  - What maximises user satisfaction
- Users never make errors in this sense!
  - Disfluences, etc, is just another behaviour the system should handle

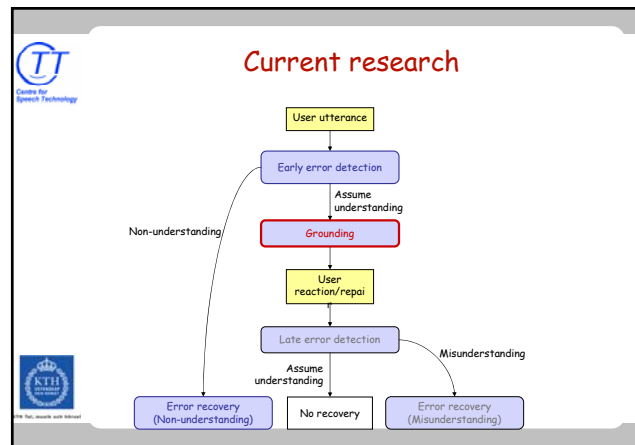
## Non-understanding error recovery

- Results show that humans tend not to signal non-understanding:
  - O: Do you see a wooden house in front of you?
  - U: YES CROSSING ADDRESS NOW  
(I pass the wooden house now)
  - O: Can you see a restaurant sign?
- This leads to
  - Increased experience of task success
  - Faster recovery from non-understanding
- Skantze, G. (2003). *Exploring human error handling strategies: implications for spoken dialogue systems.*





- ### Early error detection
- The system must understand what it doesn't understand
    - Measure of confidence in its understanding
      - Determines grounding behaviour
      - Facilitates late error detection
    - Deciding when to reject and when to accept whole utterances or parts of utterances
      - Should depend on
        - Confidence of understanding
        - Consequence of non-understanding
        - Consequence of misunderstanding



- ### Domain specific dialogue phenomena
- information seeking (adapt, etc)
  - task oriented dialogue (adapt, nice etc)
  - guiding / motivator / game leader (Higgins)
  - problem solving (trains, circuit-fix-it, nice)
  - tutoring



## Some Challenges

- Dialog Modeling
  - statistical?
- Initiative
  - conversation
- Error Handling
- Multidomain
- User modelling - Adaptivity
- Turn Taking
- Multimodal Communication