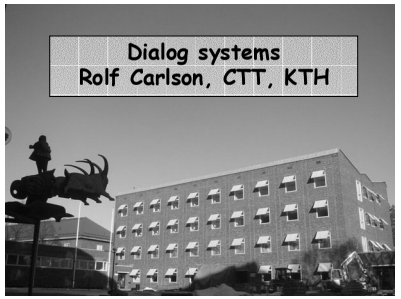
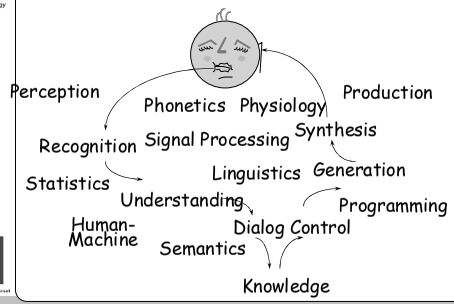


Multi Disciplinary



Agenda

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

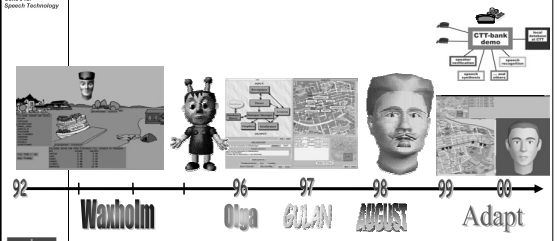
Classic systems

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

Nordic Scene

- Stockholm, Sweden
  - Waxholm
- Linköping, Sweden
  - LINLIN
- Göteborg, Sweden
  - TRINDI
- Aalborg, Denmark
- Helsinki, Finland
- Trondheim, Norway

Dialog systems at KTH



Dialog systems at KTH

2003

Dialog systems at KTH

2003

Dialog systems at KTH

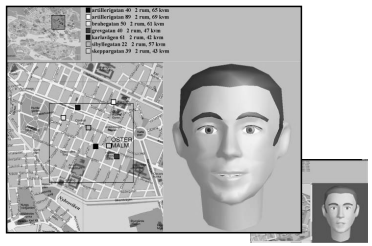
2003

AdApt multimodal dialog system

Multimodal dialog system

Conversations about appartments for sale

Work together with a animated agent, Urban



92 96 97 98 99 00

Waxholm Olga GULAN AUGUST Adapt

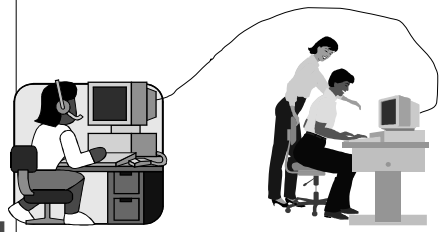
Some research issues

- Multimodal dialog modelling
  - Speech Synthesis, Animation, Turntaking
  - Speech Recognition, Pointing
  - Reference Handling
- Error Handling
- Adaptivity

Dialog Phenomena

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

Simulation (Wizard-of-Oz)



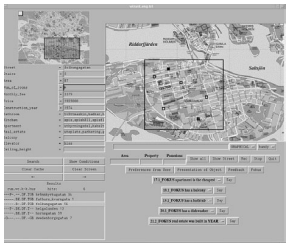
User

Human operator

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

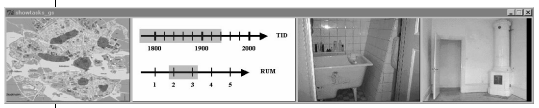
Wizard-of-Oz data collection



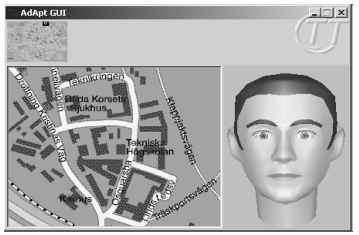
The Wizard's graphical interface

Early demo

Pictorial scenarios



Adapt – demonstration of "complete" system



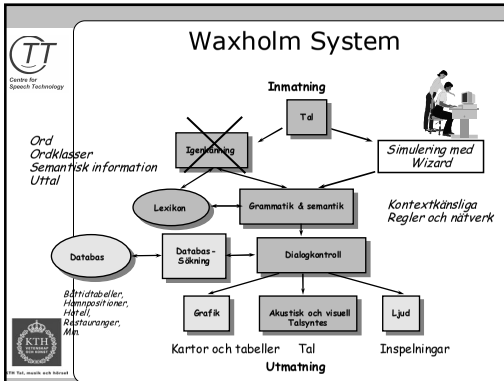
The Waxholm Project

- 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.

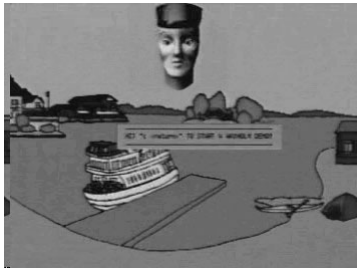
92 96 97 98 99 00

Waxholm

Waxholm System



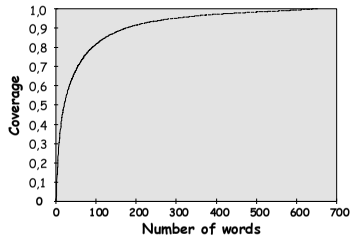
The Waxholm interface



Waxholm Database

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

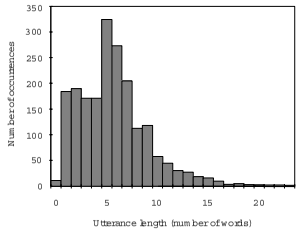
Word coverage



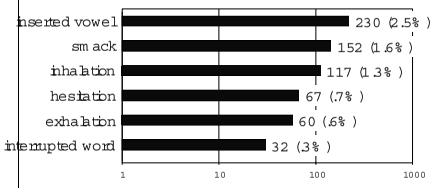
Lexicon - transcription

W ord	freq.	W ord	freq.
skÅrgÅrden	69	skulle	33
SJÖÅ3RGÅ 2N	15	SKk"ULE0	26
SJÖÅ3RGgÅ 2N	6	SKk"U	3
SJÖÅ3GgÅ 2N	5	SKk"UL	2
SJÖÅ3R#gÅ 2D dE0N	5	SKkLE0	1
SJÖÅ3RGgÅ 2D dE0N	4	SKK"UE0	1
SJÖÅ3RGgÅ 2D E0N	4		
SJÖÅ3R#gÅ 2D E0N	4		
SJÖÅ3GgÅ 2D E02N	3		
2SÅ3RGgÅ 2N	3		
SJÖÅ3RGgÅ N	2		
SJÖÅ3R#gÅ 2D dE02N	2		
SJÖÅ3R#GÅ 2D E02N	2		
SJÖÅ3#gÅ 2D dE0N	2		
SJÖÅ3RÅNG	1		
SJÖÅ3RGÅ N	1		
SJÖÅ3RGÅ 2D dE02N	1		
SJÖÅ3RGÅ 2D E0N	1		
SJÖÅ3RGÅ 2D E02N	1		
SJÖÅ3GgÅ 2D dE0N v	1		
SJÖÅ3GgÅ 2D E0N	1		
SJÖÅ3GÅ 2DE0N	1		
SJÖÅ3R#gÅ RDdE0N	1		

Utterance length distribution



Extra linguistic sounds



Number of occurrences in the Waxholm database (9200 words)

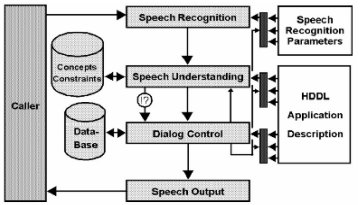
Three years later….

The Waxholm system

Instead of WOZ
”Boostrap” the system

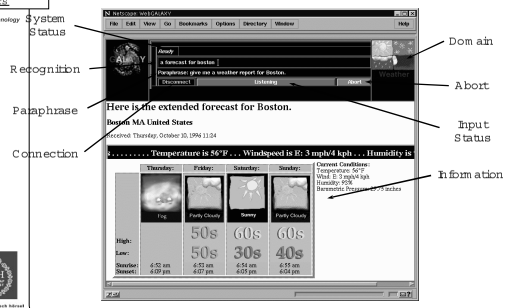
- Make a simple but complete system and evaluate
- Spread the information….
- Collect data
- Upgrade the system

Philips train information

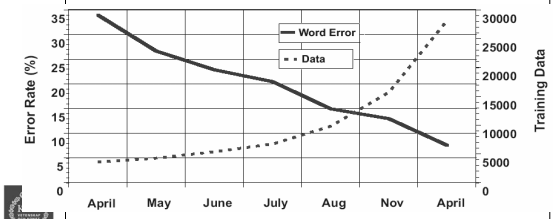


SJ 020 757575

WebGALAXY Display



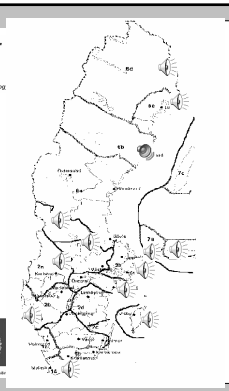
Training of Jupiter



Even generic databases are important

Swedia
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

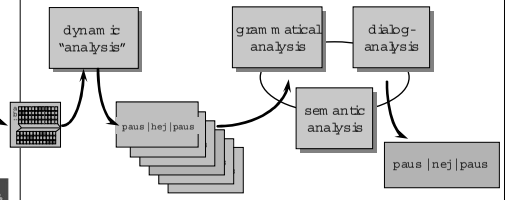
Perplexity of the language

$$H = -\sum_{\forall W} P(W)^2 \log P(W)$$

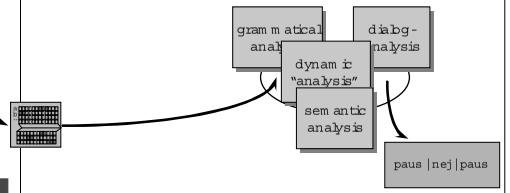
$$B = 2^H$$

B perplexity for the application
H entropy for the application
P(W) probability of a word given its preceeding context

Automatic recognition



Automatic understanding

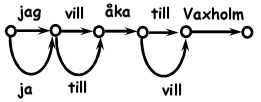


Representing multiple hypotheses

N-best list

Jag vill åka till Vaxholm
Jag till åka till Vaxholm
Jag vill åka vill Vaxholm
Jag till åka vill Vaxholm
Ja vill åka till Vaxholm
Ja till åka till Vaxholm
Ja vill åka vill Vaxholm
Ja till åka vill Vaxholm

Word graph

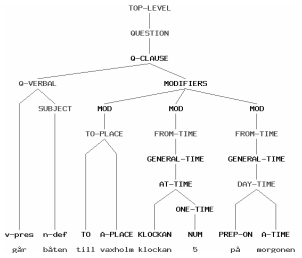


Knowledge sources - Evaluation

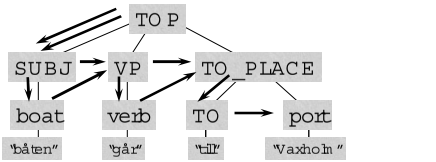
Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

→ Confidence
Expectation
Filter

Syntactic analysis



Multi-level analysis



p(boat | "TOP+SUBJ")
p(verb | "SUBJ+boat+VP")
p(TO | "VP+verb+TO_PLACE")
p(port | "VP+verb+TO_PLACE")

p(SUBJ | "TOP")
p(VP | "TOP+SUBJ")
p(TO_PLACE | "TOP+SUBJ+VP")
p(END | "SUBJ+VP+ TO_PLACE")

$$SCORE = \Sigma p_{node} + \Sigma p_{terminal} + \Sigma p_{word} + f(length)$$

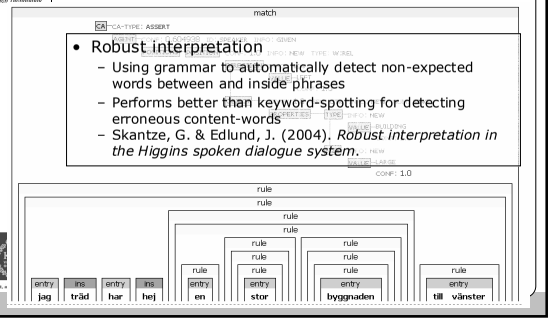
I want to go.....

Parser score

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

Robust Analysis

- 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.*



Disfluencies

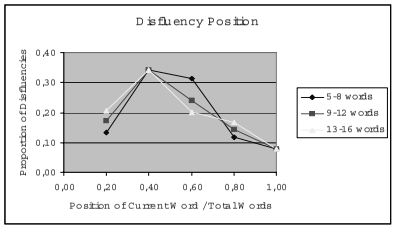
Type	Example
Filled pause	jag hum tycker om glass
Repetition	jag jag tycker om glass
Insertion	jag tycker om inte om glass
Restart	kan du jag tycker om glass
Substitution	vi tycker jag tycker om glass

'Disfluency rate'

S.Oviatt

- **Human - Human** — High
  - Two person telephone
  - Two person direct
  - One person
- **Human - Machine** — Low
  - Computer interaction

Distribution of Disfluenices



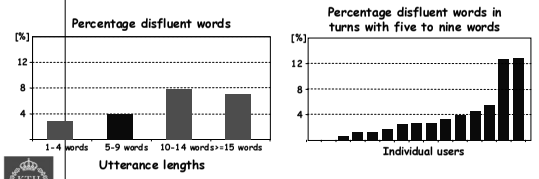
Switch board data, Liz Shriberg, Thesis, SRI

Disfluency examples from Adapt

rättelse	det är lite för ... lite för sent **tidigt** finns det nån **ehm** ... liknande ligh ... **in/-** området med som är byggd på 1800talet
avbrutet	hur **se/ eh** ... är kölet **eh** .. utrustat
pauser	**uhm** ... högt till tak och **eh** ... konstke någon kakelugn ... och balkong gärna åt söderläge
feluttal	är den eh **ny rede-** nyrenoverad
förlängning	**huuurrr** ser gatan ut
rättelseterm	jag vill gärna ha en lägenhet med ... utsiikt ... **nej** med balkong

Disfluencies in half of the Adapt corpus

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



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)

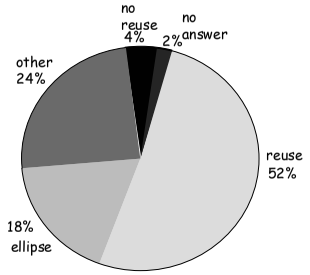
Pairs of alternative main verbs

	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)

Example of questions and answers

Hur ofta åker du utom lands på semestern?	Hur ofta reser du utom lands på semestern?
jag åker en gång om året kanske	jag reser en gång om året utom lands
jag åker ganska sällan utom lands på semester	jag reser inte ofta utom lands på semesterdet blir nära i arbetet
jag åker nästan aldrig utom lands under min semester	jag reser reser utom lands på semestern vartannat år
jag åker ungefär 2 gånger per år utom lands på semester	jag reser utom lands en gång per semester
jag åker utom lands nästan varje år	jag reser utom lands på semesterungefär en gång per år
jag åker utom lands på semestern varje år	jag brukar resa utom lands på semestern åtminstone en gång i året
jag åker utom lands ungefär en gång om året	
jag åker nästan aldrig utom lands	en gång per år kanske
en eller två gånger om året	en gång vartannat år
en gång per semester	varje år
kanske en gång per år	vartredje år ungefär
ungefär en gång per år	nu för tiden inte så ofta
Åtminstone en gång om året	varje år brukar jag åka utom lands
Varje år	

Results



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

Dialog Model

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?**

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

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

Dialog model

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

voiceXML

voicexml FORUM Sponsors and Promoters

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

http://www.voicexml.org/

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

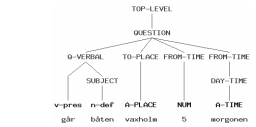
Waxholm Topics

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

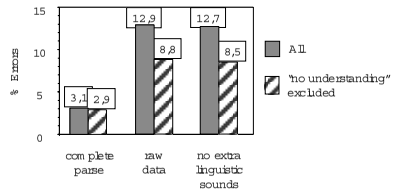
Topic selection

FEATURES → TOPIC EXAMPLES →

	TIME TABLE	SHOW MAP	FACILITY	NO UNDER-STANDING	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	.312	.031	.024	.091	.067	.091
PLACE	.091	.200	.500	.091	.067	.091
OOD	.062	.031	.122	.091	.933	.091
END	.062	.031	.024	.091	.067	.909
HOTEL	.062	.031	.488	.091	.067	.091
HOSTEL	.062	.031	.122	.091	.067	.091
ISLAND	.333	.556	.062	.091	.067	.091
PORT	.125	.750	.244	.091	.067	.091
MOVE	.875	.031	.098	.091	.067	.091

$$\underset{i}{\text{argmax}} \{ p(t_i \mid \vec{F}) \}$$

Topic prediction results



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.

Human-human conversations

	Customer		Agent	
Act	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

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

Conversational "grunts"

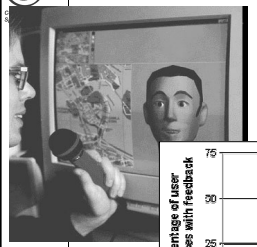
- Grunts occur an average of once every 5 seconds in American English conversation. (Nigel Ward, 2000)
- In Switchboard database
  - um was the 6th most frequent item (after I, and, the, you, and a), (Nigel Ward, 2000)
  - the four items uh, uh-huh and um and um-hum accounted for 4% of the total (Pirone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

Notation

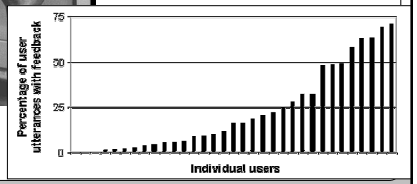
abbreviation and function/position

back	back-channel
fil	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

User studies



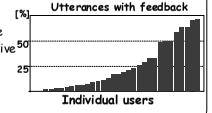
- Turn-taking
- Interaction
- Positive and Negative User Feedback
- User reactions



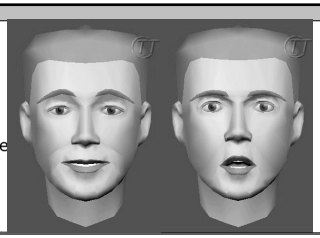
Positive and negative feedback

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



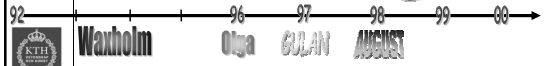
Parameter settings to create different stimuli



	Affirmative setting	Negative setting
Smile	Head smiles	Head has neutral expression
Head movement	Head nods	Head leans back
Eyebrows	Eyebrow s 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

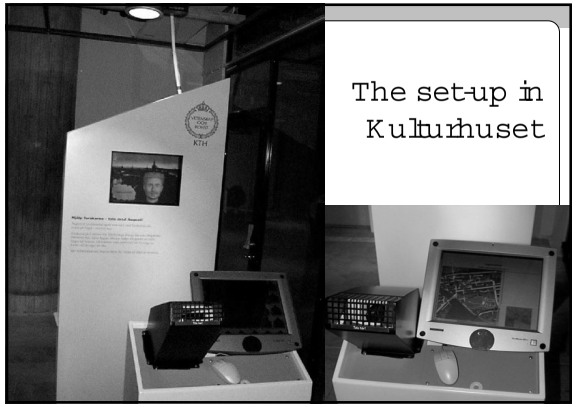
- Stockholm (events and general information)
- Yellow pages
- KTH and speech technology
- August Strindberg
- Greetings and social utterances
- Comments about the system capabilities and the discourse



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: mariatorget
- Trained on tagged N-best lists and lexicon

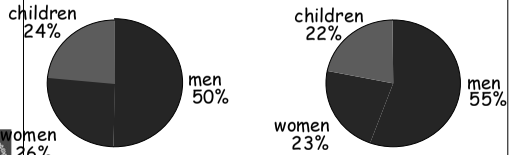
The set-up in Kulturhuset



A sample video of the system environment

The August database

September 1998 - February 1999:
10,058 utterances (approximately **15 hours** of speech) were manually checked, transcribed and analyzed



2685 users **10,058 utterances**

What do you say to August?

- Child
- Woman 1
- Woman 2

Utterance types in the August database

Socializing	Info-seeking
Social	Domain
Insult	Meta
Test	Facts

Socializing categories

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

The info-seeking categories

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

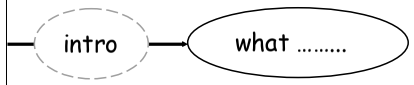
User utterance categories during the first six dialogue turns

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%

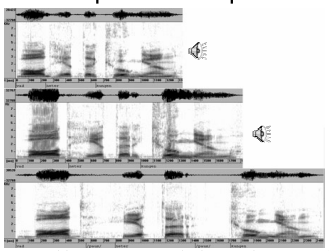
The statistics are based on the first utterances (up to six) from all users that said more than two utterances to the system

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

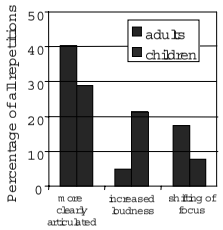


An example of a repetitive sequence



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



Some lessons for recognition

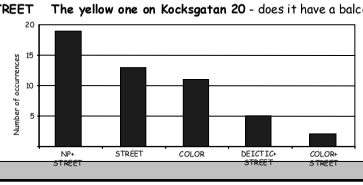
- lexical entrainment
  - use both user input and system output
- adaptive to
  - application
  - user
  - dialog
- use three recognition systems in parallel
  - continuous speech (default)
  - word by word (error resolution)
  - continuous syllables (confidence)

The August system

Verbal topicalization

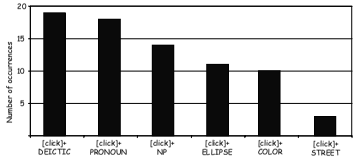
About 10% of all requests for information about a specific apartment contained a topicalized reference. 4% contained a verbal topical reference and 6% a preceding graphical reference

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

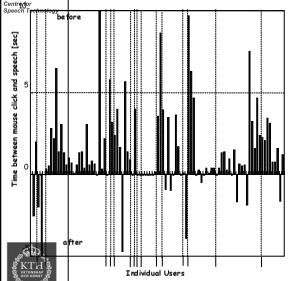


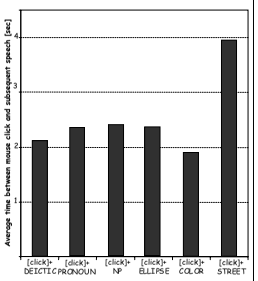
Graphical topicalization

[click] + DEICTIC	When was **this one** built
[click] + PRONOUN	Does **it** have a tiled stove
[click] + NP	Does **the apartment** have a bathtub
[click] + ELLIPSE	Which floor
[click] + COLOR	When was **the white house** built
[click] + STREET	Does **Hornsgatan 59** have a tiled stove



Timing of mouse clicks



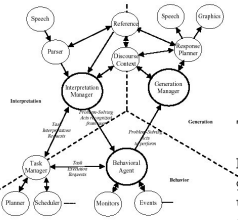


Classic systems

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

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

User and system model

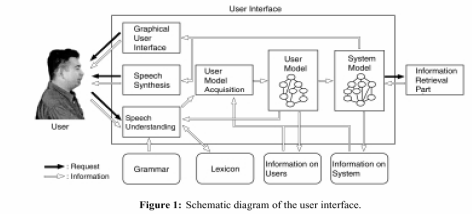
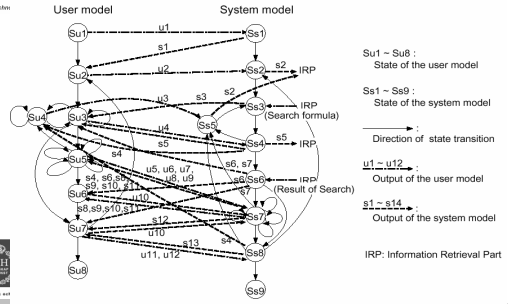


Figure 1: Schematic diagram of the user interface.

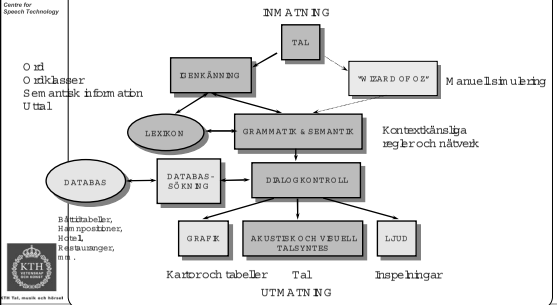
A new method for dialogue management in an intelligent system for information retrieval - *Kenji Abe, Kazushige Kurokawa, Kazunari Taketa, Sumio Ohno and Hiroya Fujisaki, ICSLP 2000*

User and system model

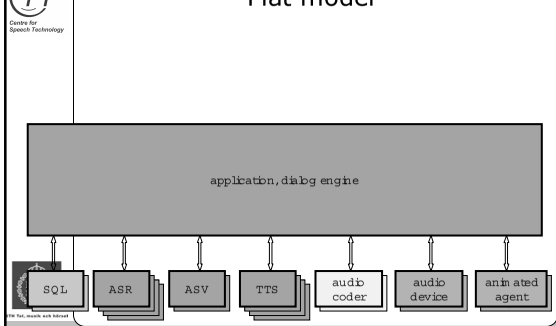


Platforms

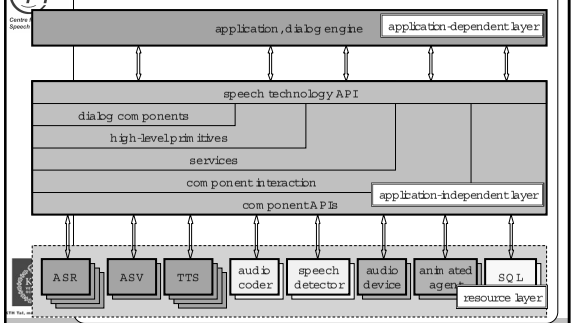
Waxholm System



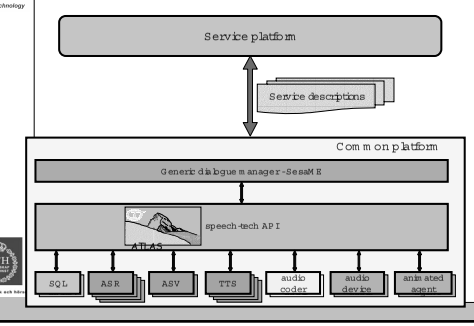
Flat model

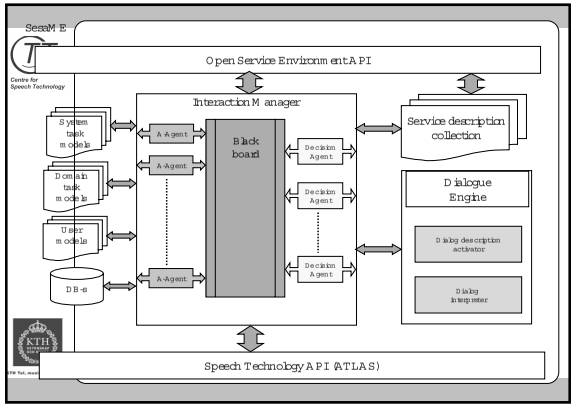


Multi-layer model



Platform support





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-)

Word accuracy

$$G = 100 * \frac{N - F - B - I}{N}$$

G	word accuracy
N	number of words
F	number of wrong words
B	number of missing words
I	number of inserted words

Not sensitive to word similarities
Ex: *i kväll - ikväll, jag - ja, Vaxholm - Vaxholms, bil - restaurang* equally wrong

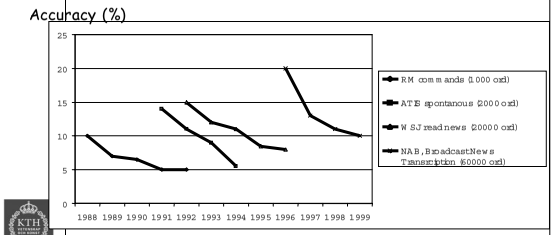
Perplexity

$$H = -\sum_{\forall W} P(W)^2 \log P(W)$$

$$B = 2^H$$

B perplexity
H entropy
P(W) probability of the word sequence in the used language Example:
Number sequences B = 11, if all digits equally probable

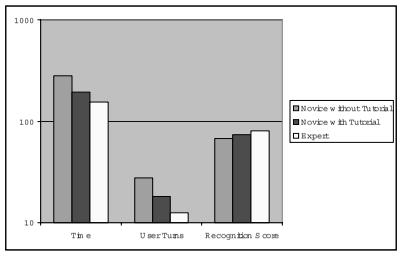
DARPA-evaluation 1988-1999



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

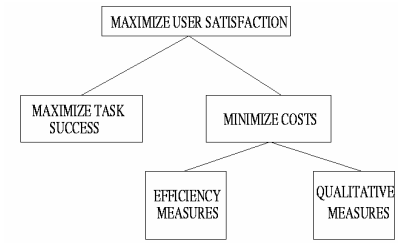
User profile



Email retrieval
Kamm, Litman Walker ICSLP98

General Models of Usability with PARADISE

Marilyn Walker, Candace Kamm and Diane Litman



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

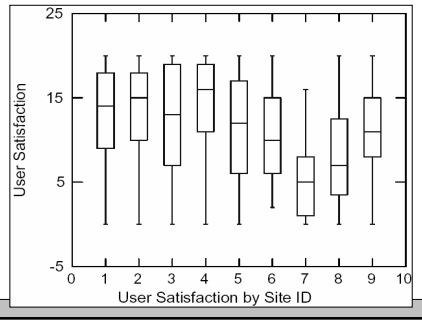
Evaluation metrics

- 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

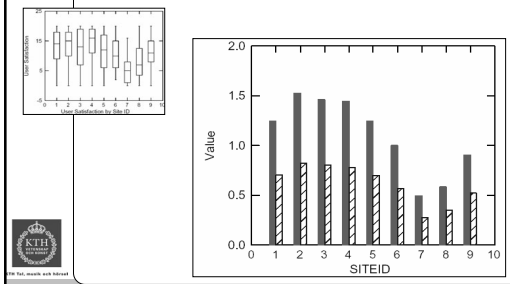
Evaluation metrics

- 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.

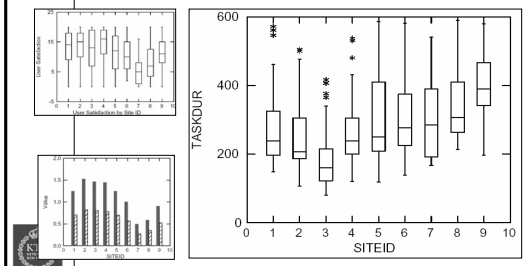
Communicator: User Satisfaction



Task completion



Task duration



Prediction of satisfaction?

PERFORMANCE = .25 MRS + .33 COMP - .33 HELP

MRS = mean recognition score
COMP = perceived completion
HELP = number of help messages
PERFORMANCE = User satisfaction

Covers 41.3% of the variance

Some Challenges

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

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

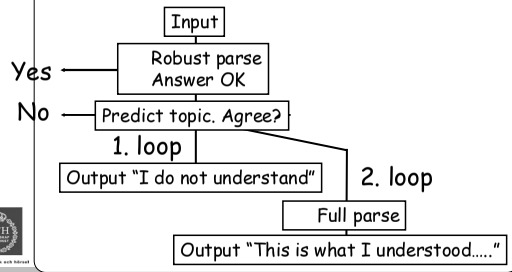
Initiative - Cue detection

- 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

Error handling in Waxholm



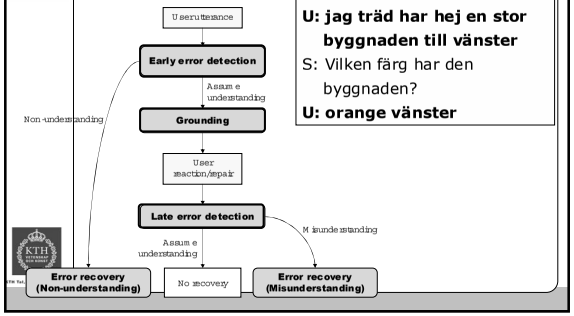
The HIGGINS domain

The HIGGINS domain

Errors in spoken dialogue systems

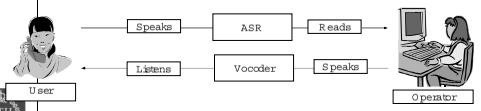
- What is an error?
  - A deviation from an expected output from a system, module or process
- Deviation from what?
  - What is written in the requirement specification
  - What a human "wizard" would do
  - What maximises user satisfaction
- Users never make errors in this sense!
  - Disfluences, etc,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

Error handling research issues



Initial experiments

- Studies on human-human conversation
- The Higgins domain (similar to Map Task)
- Using ASR in one direction to elicit error handling behaviour



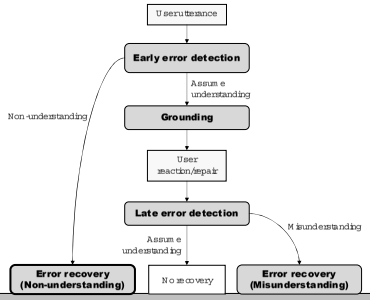
Non-understanding error recovery

- Results show that humans tend not to signal non-understanding:

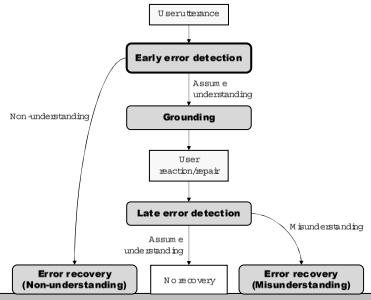
O: Do you see a wooden house in front of you?
U: **YES CROSSING ADDRESS NOW**
(I pass the wooden house now)
O: Can you see a restaurant sign?

- 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.*

Error handling research issues



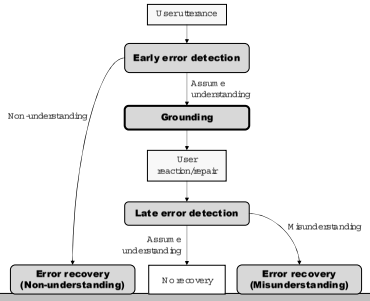
Error handling research issues



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

Current research



Higgins domain -demonstration

CHIL

"Computers in the Human interaction Loop"

- Integrated Project under the European Commission's Sixth Framework Programme.
- Coordinated by Universität Karlsruhe (TH) and the Fraunhofer Institute IITB.
- CHIL was launched on January, 1st 2004.

http://chil.server.de/

- DaimlerChrysler AG, Group Dialogue Systems, Germany
- ELDA, Evaluations and Language resources Distribution Agency, France
- IBM Ceska Republika, Jzech Republic
- RESIT, Research and Education Society in Information Technologies, Greece
- INRIA (Institut National de Recherche en Informatique et en Automatique), Project GRAVIR, France
- IRST (Instituto Trentino di Cultura), Italy
- KTH (Kungl Tekniska Högskolan), Sweden.
- CNRS, LIMSI (Centre National de la Recherche Scientifique through its Laboratoire d'Informatique pour la mécanique et les sciences de l'ingénieur), France
- TUE (Technische Universiteit Eindhoven), The Netherlands
- IPD, Universität Karlsruhe (TH) through its Institute IPD, Germany
- UPC, Universitat Politècnica de Catalunya, Spain
- Universität Karlsruhe (TH), Interactive Systems Labs, Germany
- Fraunhofer Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe, Germany
- Stanford University, Clifford Nass, USA
- CMU, Carnegie Mellon University, USA

Challenge

- The objective
  - to create environments in which computers serve humans who focus on interacting with other humans as opposed to having to attend to and being preoccupied with the machines themselves.
  - Instead of computers operating in an isolated manner, and Humans in the loop of computers we will put Computers in the Human Interaction Loop (CHIL).
- Computer Services
  - models of humans and the state of their activities and intentions. Based on the understanding of the human perceptual context, CHIL computers are enabled to provide helpful assistance implicitly, requiring a minimum of human attention or interruptions

CHIL - Services

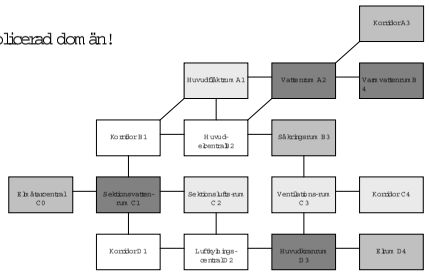
- Memory Jog (MJ).
  - It helps the attendees by providing information related to the development of the event (meeting/lecture) and to the participants. MJ provides context- and content-aware information pull and push, both personalized and public.
- Attention Cockpit (AC).
  - AC monitors the attention and interest level of participants, supporting individuals who want more or less involvement in the discussion. It can also inform the Socially-Supportive Workspaces about the attentional state of the participants.
- Connector (Connector).
  - Context-aware connecting services ensure that two parties are connected with each other at the right place, time and by the best media, when it is most appropriate and desirable for both parties to be connected.

Ljuddesign för talgränssnitt

- Utveckling av ett system för jourhavande fastighetsskötare
- Hur ser en fastighetsskötardomän ut?
  - Komplicerad! - Källare med många rum
    - vattenkranar
    - elskåp
    - Ventilationsanläggningar

Ljuddesign för talgränssnitt

Komplicerad domän!



Ljuddesign för talgränssnitt

- Systemet är användarens ”ögon och händer”, t.ex.:

S> Du befinner dig i korridor C4. Härifrån kan du gå till följande rum: elrum D3, sektionsvattenrum C3. Vad vill du göra?
A> gå till sektionsvattenrum C3
S> Du befinner dig i sektionsvattenrum C3. Härifrån kan du gå till följande rum: korridor C4, elmåtarcentral C2, ventilationsrum E3. Vad vill du göra?
A> titta på vattenkranarna
S> Du tittar på sektionsvattenkranarna. Följande kranar är öppna: kran 2, kran 7. Resten av kranarna är stängda. Vad vill du göra?

Ljuddesign för talgränssnitt

- Hur kan ljud användas som navigationsstöd?
- Auditiva ikoner
  - två typer förekommer i två olika roller
    - bakgrundsljud (kontextgivare)
    - återkoppling på kommando
  - [exempeldialog utan ljud]
  - [exempeldialog med ljud]

The End