**Dialog systems at KTH**

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

**Multi Disciplinary**
- Perception
- Phonetics
- Physiology
- Production
- Recognition
- Speech Processing
- Synthesis
- Statistics
- Linguistics
- Generation
- Understanding
- Programming
- Dialog Control
- Knowledge
- Human-Machine
- Semantics

**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

The HIGGINS domain

AdApt multimodal dialog system

Conversations about apartments for sale

Work together with an animated agent, Urban

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)

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

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.

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

Coverage

0
100
200
300
400
500
600
700

Number of words

0
0,1
0,2 0,3
0,4
0,5
0,6 0,7 0,8 0,9
1,0

Word coverage

Lexicon - transcription

Word freq.

skärgården 69
skulle3 3
SJ'ÄÄRGÅ:2N 15
SKk"ULE0 26
SJ'ÄÄRGÅ:2N 6
SKk"UL 2
SJ'ÄÄRGgÅ:2DN 5
SKk"UL 2
SJ"ÄÄRg`Å:2DE0N 5
SKkLE0 1
SJ"ÄÄRg`Å:2DE0N 4
SJ"ÄÄRÅ:NG 1
SJ"ÄÄRGÅ:N1
SJ"ÄÄRGÅ:2DdE02N 1
SJ"ÄÄRGÅ:2DE0N 1
SJ"ÄÄRGÅ:2DE02N 1
SJ'ÄÄGgÅ:2DdE0N 1
SJ'ÄÄGgÅ:2DE0N 1
SJ'ÄÄGÅ:2DE0N 1
SJ"ÄÄRg`Å:RDdE0N 1
Utterance length distribution

Extra linguistic sounds

<table>
<thead>
<tr>
<th>Sound</th>
<th>Number of Occurrences</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inserted vowel</td>
<td>230</td>
<td>2.5%</td>
</tr>
<tr>
<td>Sniff</td>
<td>393</td>
<td>4.4%</td>
</tr>
<tr>
<td>Exhalation</td>
<td>117</td>
<td>1.3%</td>
</tr>
<tr>
<td>Inhale</td>
<td>67</td>
<td>0.7%</td>
</tr>
<tr>
<td>Exhale</td>
<td>60</td>
<td>0.7%</td>
</tr>
<tr>
<td>Interrupted word</td>
<td>33</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Number of occurrences in the Waxholm database (9200 words)

Three years later....

The Waxholm system

Instead of WOZ
“Bootstrap” the system

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

Waxholm is shown on this map

I think I want to go to Waxholm

I am looking for boats to Waxholm

From where do you want to go

The city

Which day of the week do you want to go?

I want to go tomorrow

There are lots of boats from Stockholm to Waxholm on a Friday

At what time do you want to go?

When do the evening boats depart?

This is a table of the boats...

Is it possible to eat in Waxholm?

Information about the restaurants in Waxholm is shown in this table

Philips train information

020 757575
Swedish dialects

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

- Född i USA
- ex Jugoslavien

Even generic databases are important

Swedia
SpeechDat

Speech understanding some aspects

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

Perplexity of the language

\[ H = - \sum_{W} P(W) \log P(W) \]

\[ B = 2^H \]

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

Automatic understanding

Representing multiple hypotheses

Knowledge sources - Evaluation

Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

Confidence
Expectation
Filter

Syntactic analysis

Multi-level analysis

N-best list
Word graph

Knowledge sources - Evaluation

Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

Confidence
Expectation
Filter

Syntactic analysis

Multi-level analysis

Knowledge sources - Evaluation

Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

Confidence
Expectation
Filter

Syntactic analysis

Multi-level analysis

Knowledge sources - Evaluation

Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

Confidence
Expectation
Filter

Syntactic analysis

Multi-level analysis

Knowledge sources - Evaluation

Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

Confidence
Expectation
Filter

Syntactic analysis

Multi-level analysis

Knowledge sources - Evaluation

Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

Confidence
Expectation
Filter

Syntactic analysis

Multi-level analysis

Knowledge sources - Evaluation

Acoustic analysis
Syntactic analysis
Semantic analysis
Dialog state
Dialog Context

Confidence
Expectation
Filter

Syntactic analysis

Multi-level analysis
I want to go......

12.26 Jag vill låka från Stockholm till Vaxholm.
10.01 Jag vill låka till Vaxholm.
09.85 Jag skulle åka till Vaxholm.
05.30 Jag vill låka.
03.17 Närgårdeten båttill Vaxholm?
-1.32 Närgårdeten båttill Vaxholm?
00.00 Jag vill låka till mitt omkring.

Disfluencies

- Robust interpretation
  - Using grammar to automatically detect non-expected words between and inside phrases
  - Performs better than keyword-spotting for detecting erroneous content-words

Human - Human
- Two person telephone: Highly
- One person: Low

Human - Machine
- Computer interaction: Low

'Sdisfluency rate'

Distribution of Disfluences

Disfluency examples from Adapt

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filled pause</td>
<td>jag tycker om glass</td>
</tr>
<tr>
<td>Repeition</td>
<td>jag tycker om glass</td>
</tr>
<tr>
<td>Insertion</td>
<td>jag tycker om glass</td>
</tr>
<tr>
<td>Restart</td>
<td>jag tycker om glass</td>
</tr>
<tr>
<td>Substitution</td>
<td>vilken jag tycker om glass</td>
</tr>
</tbody>
</table>
Disfluencies in half of the Adapt corpus

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

Percentage disfluent words
Percentage disfluent words in turns with five to nine words

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

Utterance Generation

Percentage disfluent words

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)

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)

Pairs of alternative main verbs

- höra-lyssna (listen - hear)
- sevä-på (watch- go to)
- köpa-handla (shop-buy)
- föredrar-tycker mest om (prefer -like the most)
- testa-pröva (test-try)
Exemple of questions and answers

Human-machine interaction
- Initiative
- Who is the user
- First time?
- Terminology
- 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

Results
- No reuse
- 4%
- Reuse
- 52%
- Ellipse
- 18%
- Other
- 24%
- No answer
- 2%
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

- Spoken, –4
  - NCC, Now and Molendo
- Promoters –25
  - Nuance Communications Inc.
  - Amgen
  - Angiodynamics Inc.
  - Bionica, Inc.
  - Beryl, Inc.
  - ComWin
  - Corning Network Systems
  - Covia
  - Digital Network Systems
  - Harp, Ltd.
  - Control Research Lab
  - Honeywell
  - InterAct Systems
  - Lexicon & Haapio
  - Micromedia
  - MedianWeb
  - MindWave

Dialog 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

SpeechObjects™ & VoiceXML

- Phone: Certain personal information can be shared with different users
- Reliable: A person can use voice input to access voice data
- Natural: The user can use voice data to get back to the order of voice data
- Secure: The user can use voice data to get back to the order of voice data
- Mobile: The user can use voice data to get back to the order of voice data
- Encryption: The user can use voice data to get back to the order of voice data
- Nuance Voyager
- Waxholm Topics
  - TIME_TABLE Task: get a time-table.
  - SHOW_MAP Task: get a chart or a map displayed.
  - EXIST Task: display lodging and dining possibilities.
  - OUT_OF_DOMAIN Task: the subject is out of the domain.
  - NO_UNDERSTANDING Task: no understanding of user intentions.
  - END_SCENARIO Task: end a dialog.
Semantic Frame

Current functions: /TO-PLACEQ_VERBALSUBJECTFROM-TIME/
Current meaning: /MOVEBOATPORTQUANT/

History functions: /TO-PLACEQ_VERBALSUBJECTFROM-TIME/
History meaning: /MOVEBOATPORTQUANT/

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

TIME SHOW FACILITY NO UNDER-OUT OF END

TABLE MAP STANDING DOMAIN

OBJECT 0.062 0.312 0.073 0.091 0.067 0.091
QUEST-WHEN 0.188 0.031 0.024 0.091 0.067 0.091
QUEST-WHERE 0.062 0.688 0.390 0.091 0.067 0.091
FROM-PLACE 0.250 0.031 0.024 0.091 0.067 0.091
AT-PLACE 0.062 0.219 0.293 0.091 0.067 0.091
TIME 0.312 0.031 0.024 0.091 0.067 0.091
PLACE 0.091 0.200 0.500 0.091 0.067 0.091
OOD 0.062 0.031 0.122 0.091 0.933 0.091
END 0.062 0.031 0.024 0.091 0.067 0.909
HOTEL 0.062 0.031 0.488 0.091 0.067 0.091
HOSTEL 0.062 0.031 0.122 0.091 0.067 0.091
ISLAND 0.333 0.556 0.062 0.091 0.067 0.091
PORT 0.125 0.750 0.244 0.091 0.067 0.091
MOVE 0.875 0.031 0.098 0.091 0.067 0.091

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

<table>
<thead>
<tr>
<th>Custom</th>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledge</td>
<td>47.9 2.3 30.8 3.1</td>
</tr>
<tr>
<td>Request</td>
<td>29.5 9.0 15.9 12.9</td>
</tr>
<tr>
<td>Content</td>
<td>13.1 5.3 11.3 6.4</td>
</tr>
<tr>
<td>Item</td>
<td>5.8 7.3 27.8 12.7</td>
</tr>
<tr>
<td>Item unit</td>
<td>3.4 6.9 15.1 6.7</td>
</tr>
</tbody>
</table>

Conversational “grunts”

- Grunts occur an average of once every 5 seconds in an edition of YLI conversation. (Hill, 2001)
- A Switchboard database
  - on, the, other, and, from, the
  - and, you, and, with, amounted to the total
  - of these, and, 1991.

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

abbreviation and function/position

**back** back-channel
**fill** filler, including various things that occur utterance-

on turn-initially
**dis** disfluency utter
**is** isolate, produced when neither person has the turn,

typically one self-directed than other-directed
**con** confirmation, in response to a back-channel

**oth** other, including clause-fillers, items that occur

in quotations, and items whose function is obscure

User studies

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

![Bar chart showing user feedback](chart.png)

User Feedback

<table>
<thead>
<tr>
<th>System</th>
<th>User</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>94%</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>4%</td>
<td></td>
</tr>
</tbody>
</table>

Parameter settings to create different stimuli

<table>
<thead>
<tr>
<th>Setting</th>
<th>Smiles</th>
<th>Head position</th>
<th>Eye movement</th>
<th>Eye brows</th>
<th>Eye closure</th>
<th>F0 contour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affirmative</td>
<td>Smile</td>
<td>Head nods</td>
<td>Head moves</td>
<td>Eyebrows</td>
<td>Eyesclose</td>
<td>Declarative</td>
</tr>
<tr>
<td>Negative</td>
<td>No smile</td>
<td>Head back</td>
<td>Head leans</td>
<td>Eyebrows</td>
<td>Eyesclose</td>
<td>Interrogative</td>
</tr>
</tbody>
</table>

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
  - semantic features from lexicon
  - Feature/value representation
  - Trained on tagged N-best lists and lexicon
The setup 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.

What do you say to August?

- Child
- Woman 1
- Woman 2

Utterance types in the August database

<table>
<thead>
<tr>
<th>Socialization</th>
<th>Info-seeking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td>Domain</td>
</tr>
<tr>
<td>Insult</td>
<td>Meta</td>
</tr>
<tr>
<td>Test</td>
<td>Facts</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Domain</th>
<th>How many books did Strindberg write?</th>
<th>Where are the restaurants on Kungsatan?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta</td>
<td>What can I ask you?</td>
<td>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</td>
</tr>
<tr>
<td>Facts</td>
<td>What’s the capital of Finland?</td>
<td>What is two times two? How many people live in Madrid?</td>
</tr>
</tbody>
</table>

User utterance categories during the first six dialogue turns

<table>
<thead>
<tr>
<th>category</th>
<th>children</th>
<th>women</th>
<th>men</th>
</tr>
</thead>
<tbody>
<tr>
<td>only socializing</td>
<td>34%</td>
<td>31%</td>
<td>35%</td>
</tr>
<tr>
<td>only info-seeking</td>
<td>28%</td>
<td>39%</td>
<td>34%</td>
</tr>
<tr>
<td>from socializing to info-seeking</td>
<td>37%</td>
<td>35%</td>
<td>43%</td>
</tr>
<tr>
<td>alternating</td>
<td>7%</td>
<td>6%</td>
<td>2%</td>
</tr>
</tbody>
</table>

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

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

- Clear articulation
- Increased loudness
- Shifting of focus

Percentage of all repetitions

- Adults
- Children

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)
What is your name?
I call myself Strindberg, but I don’t really have a surname

When were you born?
Strindberg was born in 1849

What do you do for a living?
I can answer questions about Strindberg, KTH and Stockholm

How many people live in Stockholm?
Over a million people live in the Stockholm area

Do you like it here?
People who live in glass houses should not throw stones
Yes, that was a smart thing!

I come from the department of Speech, Music and Hearing
The Royal Institute of Technology!

The information is shown on the map

Thank you!
You are welcome!

Good bye!
Perhaps we will meet soon again!

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.

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
  - Verbimobil

TRIPS
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.

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

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.

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

Platforms

Waxholm System

GRAMMATIK & SEMANTIK
DIALOGKONTROLL
GRAFIK
INGENARIO
AKUSTISK OCH VISUELL TALSYNTES
DATABAS-SÖKNING
"WIZARDFOZ"
Kartor och tabeller, Båttidtabeller, Hamnpositioner, Hotell, Restauranger, mm.

Inmanelling
UTMATTNING

Kontextkänsliga regler och nätverk
LEXIKON

Manuellsimulering

Inspelningar

DATABAS
ORD
ORDKASSER
SEMANTISK INFORMATION
UTTAL

TAL

Flat model

Multi-layer model

User and system model

Multi-layer model

Componet APIs

APPLICATION-DEPENDENT LAYER

APPLICATION-DEPENDENT LAYER
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)
  - Topic Detection and Tracking (1998-, English, Mandarin)
  - Information Extraction - Entity Recognition (1999-)

Word accuracy

\[ G = 100 \times \frac{N - F - B - I}{N} \]

\( G \) 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: ikväll - ikväll, jag - ja, Vaxholm - Vaxholms, bil - restaurang equally wrong
Perplexity

\[ H = - \sum_{w \in V} P(w) \log P(w) \]

B = 2^H

Perplexity

B = entropy

w probability of the word sequence in the used language

Example: Number sequences B = 11, if all digits equally probable

DARPA-evaluation 1988-1999

Accuracy (%)

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 (CARS-LIMSI), France
  - Universität Stuttgart, Germany
  - Kungliga Tekniska Högskolan (KTH), Sweden
  - Vocalis Ltd, England
  - Daimler-Chrysler AG, Germany
  - ELNET, Europe
- URL: www.disc.dk

User profile

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.
- 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
**Prediction of satisfaction?**

\[
\text{PERFORMANCE} = 25 \times \text{MRS} + 33 \times \text{COMP} - 33 \times \text{HELP}
\]

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

PERFORMANCE = User satisfaction

Covered 41.3\% of the variance

**Some Challenges**

- Dialog Modeling
- Initiative
- 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

**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

Input
Robust parse
Answer OK

Yes
No

Predict topic. Agree?
1. loop
Output "I do not understand"
2. loop
Full parse
Output "This is what I understood..."

The HIGGINS domain

The primary goals of HIGGINS are to navigate in the environment,
provide simple information about the immediate surroundings.

This is a 3D test environment

The HIGGINS domain

U: Jag har en stor byggnad till vänster
S: Vilken färg har den byggnaden?
U: Orange
S: Beskriv något mer
U: Jag har en glasbyggnad framför mig
S: Ställ dig mellan dem och en träbyggnad

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

User utterance
Operator reaction/repair
Assume understanding
Non-recovery
Non-understanding
Assume understanding
Error recovery (Non-understanding)
Early error detection
Grounding
Late error detection
Error recovery (Misunderstanding)

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


**Error handling research issues**

- **Early error detection**
  - Measure of confidence in its understanding
  - Determines grounding behaviour
  - 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
"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/

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
  - memory jog (MJ): It helps the attendees by providing information related to the development of the event (meeting/lecture) and to the participants.
  - attention cockpit (AC): It 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

Systemet är användarens "ögon och händer", t.ex.:


A> gå till sektionsvattenrum C3


A> titta på vattenkranarna


Ljuddesign för talgränssnitt

• Hur kan ljud användas som navigationsstöd?

• Auditive ikoner
  - två typer förekommer i två olika roller
    • bakgrundsljud (kontextgivare)
    • återkoppling på kommando
  - [exempeldialog utan ljud]
  - [exempeldialog med ljud]

The End