CoSy @ Onassis

H. I. Christensen

Cognitive Systems for Cognitive Assistants

Henrik I. Christensen

Centre for Autonomous Systems
Kungl Tekniska Högskolan
Stockholm, Sweden
hic@kth.se

Robotics and Intelligent Machines
Georgia Institute of Technology
Atlanta, GA 30332
hic@cc.gatech.edu

Thank you

- Honored to be part of the lecture series
- The Onassis Foundation for the invitation and for selecting robotics (and humans) as a focus for a workshop
- Takeo Kanade for the organisation of the technical content

Acknowledgment

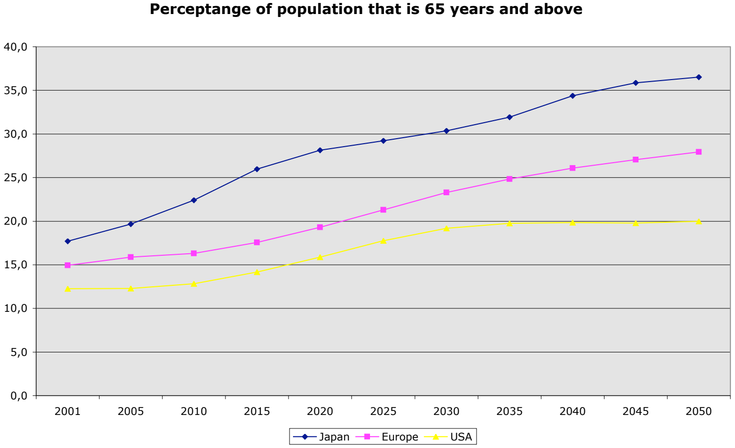
- Research involving a large number of people
- Robot navigation
  - P. Jensfelt, W. Burgard, E. A. Topp, & O. Martinez
- Spoken Dialogue
  - G-J. Kruiff
- Scene Interpretation
  - J. Wyatt, A. Sloman, & N. Hawes
- Vision and Recognition
  - A. Pronobis, B. Caputo, A. Leonardis, B. Leibe, & B. Schiele
- Human Robot Interaction
  - E. Pacchierotti, K. Severinson-Eklundh, H. Huttenrauch, & A. Green
- Funded by the European Commission as part of the Cognitive Systems Programme - Project CoSy and the Beyond Robotics Project – Cogniron

Outline

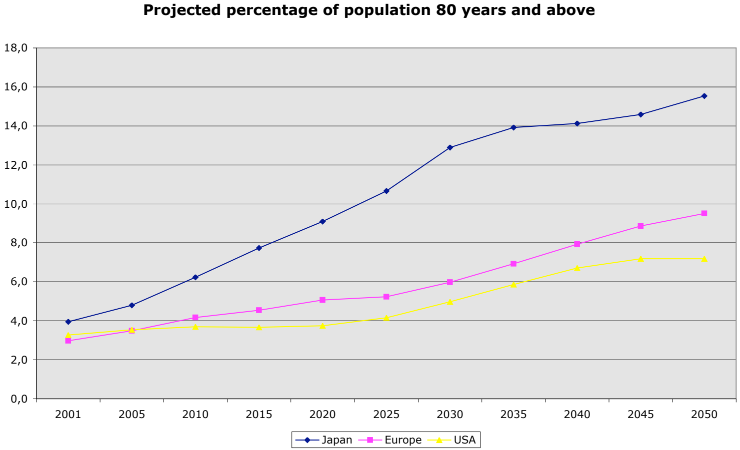
Societal Motivation

- Society is going through a significant change in demographics.
- 50%+ more people older than 60 years of age
- 100%+ more people with an age above 80.
- Will influence our society in a number of different ways

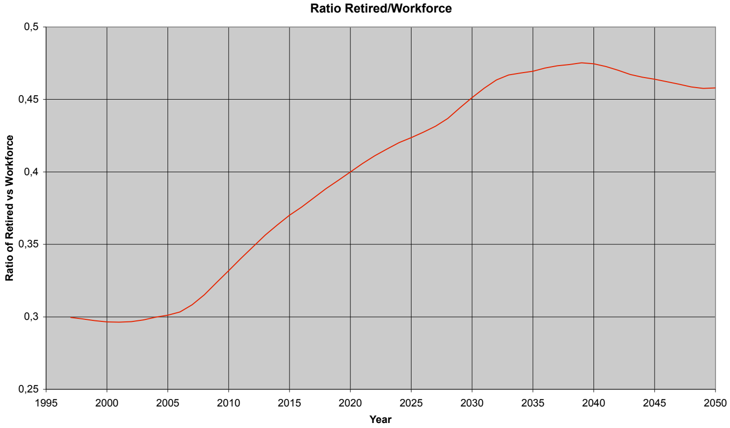
Demographics 101



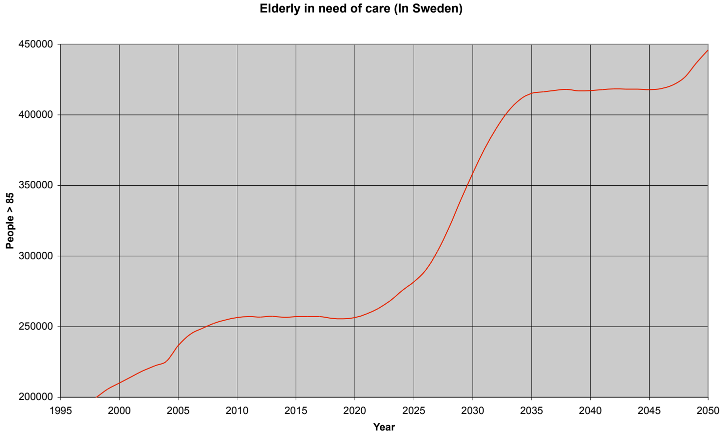
Demographics 101



Swedish Demographics 65+



Swedish Demographics 80+



Societal implications

- 200%+ increase in productivity
- Significant changes in benefits package
- Changes to social welfare support
- 50+ will become the majority
- Housing market will change
- Assistance required in the homes

Elderly Assistance

- Have a sense of autonomy
- Remain in their normal neighborhood
- Select their own clothing
- Decide on their own routines
- Perform normal tasks without a need for external assistance.
- Keep memories alive

Personal Robotics

CoSy @ Onassis

H. I. Christensen

Introduction

Early Evaluation

Explorer

Social Interaction

Perspectives

Examples

A cognitive assistant?

Studies of embodied systems for interaction with people in everyday settings

Study performed over 4 years of research

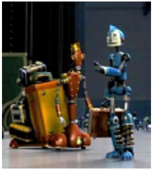
Objectives / Context

- Study of methods for (computational) cognitive systems
- An integrated approach to the study of systems in terms of
  - components such as perception, reasoning, architecture, world models, perception-action modelling ...
- Demonstration of methods in integrated systems using well defined scenarios
  - Explorer, PlayMate, & Philosopher

Objectives refined

- Two types of objectives
  - Theory and Implementation / Empirical
- Theory Objectives
  - Architecture, perception-action, communication, deliberation, reflective, affective/motivational
- Implementation Objectives
  - Integration into Systems, Nature vs. Nurture

Scenario based integration / Demonstrators



Explorer:
- Spatial exploration
- Reasoning about space
- Talking about space

PlayMate:
- Replication of manipulative actions
- Scene Interpretation
- Action Modelling & Articulation

Philosopher:
- Introspection of knowledge
- Reasoning about affordances ...
- The meaning of . . .

Research Challenges

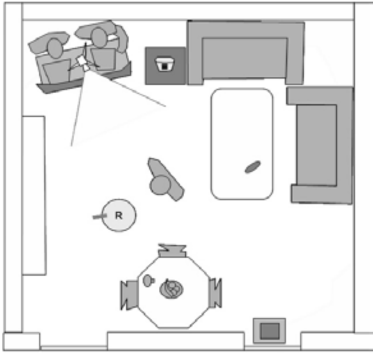
- Architecture
- Knowledge Generation
- Perception-Action Integration
- Flexible Planning and Recovery
- Flexible Interfaces
- Dealing with Novelty
- Introspection of Knowledge, Plans and Actions

Outline

What do people really need?

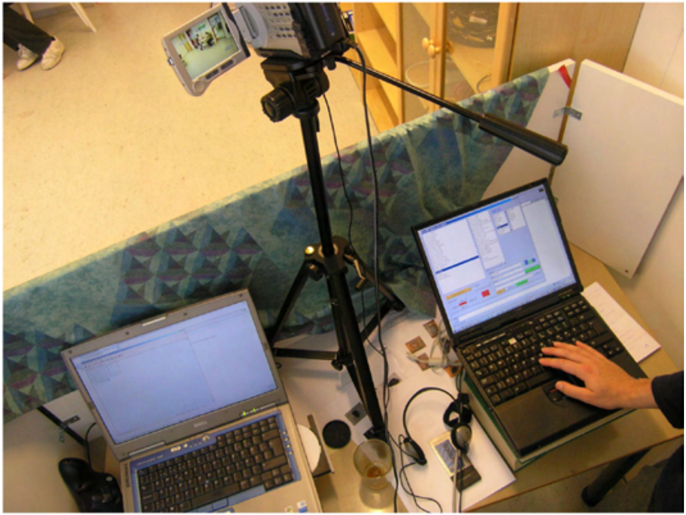
- Early evaluation of user behaviour as part of design process
- High fidelity simulation of systems - Wizard-of-Oz study to determine behaviour
- Three studies have been performed to analyze robot behaviour
    1. The HomeTour to show a robot around - Living room
    2. A tour of a laboratory to determine joint representations
    3. Social interaction with a robot - The Italian driving hypothesis?

First Wizard of Oz Study



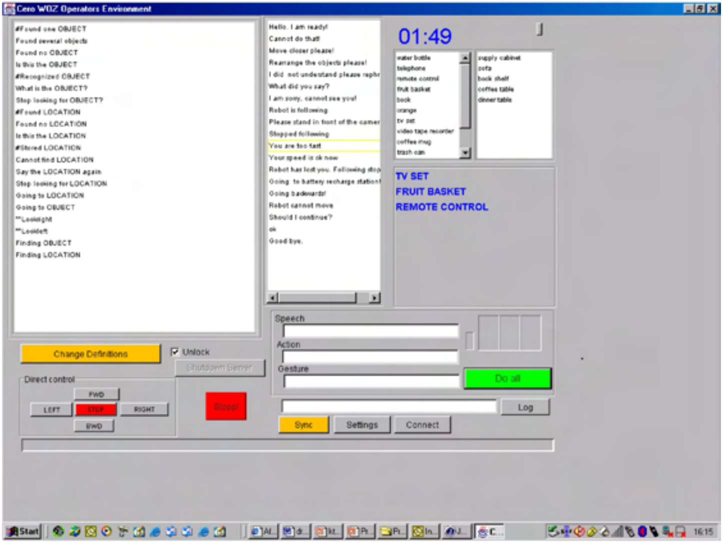
- Introduce a robot to a living room
- Identify the main objects and name them
- Questions:
  - Dialogue behaviour
  - Breakdown?
  - Spatial distances in interaction
  - Speech / Gesture usage

WoZ setup



The table for control of interaction

WoZ Speech Interface



WoZ Speech Acts

Tag	Description	Example
REJECT	Reject the proposition by the last speaker.	*R: Cannot do that*
ACK	Acknowledge the proposition by the last speaker	*R: Ok.*
DIRECTIVE	An action directive command aimed to influence the behavior of the listener.	*U: Go backwards.*
ASSERT	A proposition asserting some property.	*U: This is a chair.*
STOP	Stop is a frequently directive command used to stop the robot's movements.	*U: Stop*
FOLLOW	Follow activates the follow-behavior of the robot.	*U: Follow me*
REQ-ATT	Any utterance or gesture aimed at grabbing the attention of the robot.	*U: Hello robot!* + Wave
FB-EVAL	An utterance providing positive or negative evaluation.	*U: Good work!*

WoZ Example

WoZ Video Example

WoZ Example of Naming

Image 1-6, numbered left to right (in a 2 X 6 grid)

- Pointing gestures
- Robot can "train" users
- Shared attention is important
- People have more than expected patience

WoZ Spatial Relations Analysis

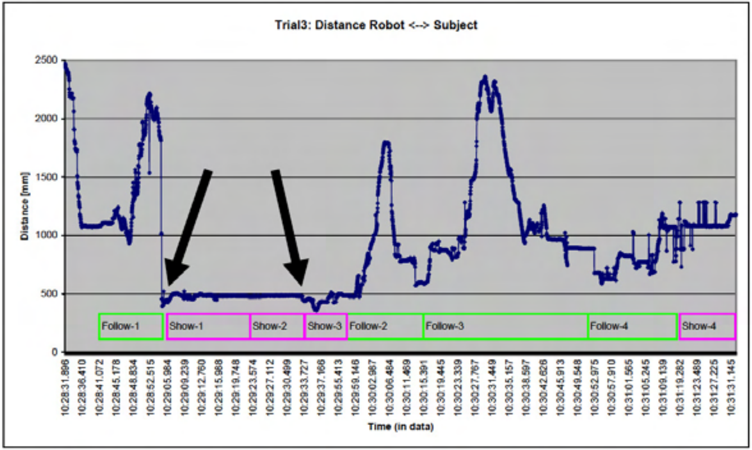


Figure 5: Robot centric laser data plot, showing distance between robot and subject during different episodes

WoZ-1 Analysis

- People are willing to accept robot behaviour
- Limited, if any, understanding of underlying system
- A clear dialogue model is needed - to avoid break down
- Robot Gaze is important to signal shared attention and as a feedback mechanism
- People have a poor understanding of the system capabilities which result in “surprising” behaviour.

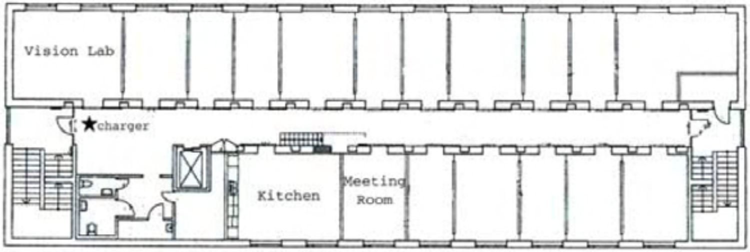
2nd WoZ Experiment

- Is a “home-tour/office-tour” person specific?
- How accurate are people in their guidance?
- How does robot behaviour influence the dialogue?
- Are there variations across “types” of people?

WoZ-2 Study

- Tour of an office complex
- Vision Researchers, Robotic Researcher and Administrators as test subjects – is there a difference?
- Task: "Show all the important objects"

WoZ-2 Test Environment



WoZ-2 Example Video

WoZ-2 Results

Table 1: Quantifiable results from the pilot study

Observation Subject	VR	VR	VR	SE	RR
Interaction time	22 min	19 min	11 min	25 min	24 min
# regions	4	2	–	2	2
# locations[I]	4	4	5	4^{II}	8^{III}
# regions w o loc.	3	2	–	1	1
# loc. w o region	3	4	5	2	3^{IV}
# regions w o entering	1	2	1	1	–
Behaviour noticed	Yes	Yes	–	No	Yes
– appropriate	Yes	Yes	–	–	Yes
– appears smart	Yes	No	–	–	Yes

VR: Vision researcher, SE: Secretary,
RR: Robotics researcher

I: including regions that were only pointed to
II: including one small object (salt)
III: including one person and two doorways
to respective rooms
IV: excluding doorways

WoZ-2 Lessons

- Spatial models are highly subjective
- (Educated) people make significant assumptions about "robot" capabilities
- Linguistic ambiguities are significant in these systems
- A pure follow / speech / gesture model is difficult to implement.
- Cognition, situation awareness, and generalization are important aspects to consider.

Questions / Comments?

"Yes ... I believe there's a question in the back."

Outline

Explorer Introduction

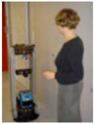
- Robots enter into our daily lives for services
- Tour guide for people in public spaces
- Fetch-and-carry for domestic/office services
- How can we endow the systems with a user representation of space?

The Easy Solution

"Now! …. That should clear up a few things around here!"

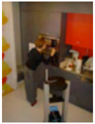
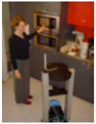
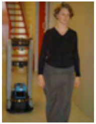
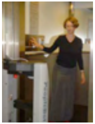
The Problem

- Mapping as done in robots is well suited for robots
- It might be a poor match for human interaction
- What do people asssociate with in navigation?
- How can human $\leftrightarrow$ robot models be reconciled?



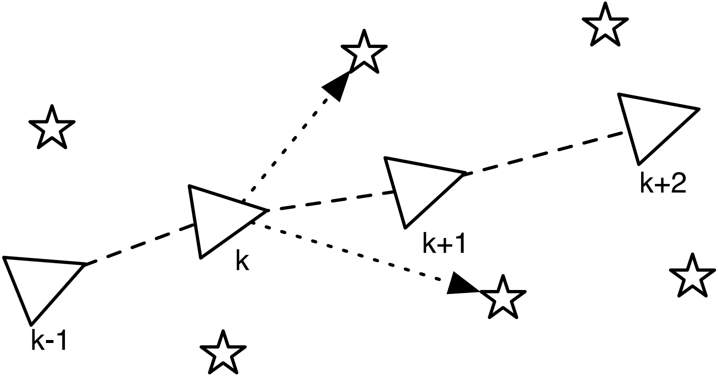


Components for the Puzzle

- A robust SLAM system
  - Integration of features, robustness,
- A dialogue system for human interaction
  - A spoken dialogue is used here
- A strategy to integrate the two models
  - Here the unification of representations is key

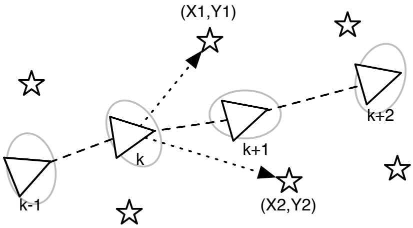
The basic problem layout



The localisation problem

Detection and estimation of position of features over time. The state is here $\vec{x}$ and range and bearing to features is measured. The pose estimate is optimized. Given a known layout of features we minimize the functional:

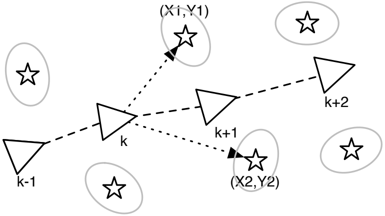
$$\vec{x} = \text{argmin}_{\vec{x}} \sum_i \sqrt{(\cos(\phi)X_i + \sin(\phi)Y_i - t_x - X_{mi})^2 + (-\sin(\phi)X_i + \cos(\phi)Y_i - t_y - Y_{mi})^2}$$



The mapping problem

Bearing and range is measured to landmarks, and the position is optimized over time. This can be done by a Kalman filter where

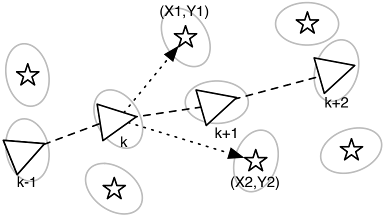
$$\vec{x}_m = \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \\ \vdots \\ X_N \\ Y_N \end{bmatrix}$$



The SLAM problem

Both robot pose and landmarks have associated uncertainty. The estimation problem is then:

$$\vec{x} = \begin{bmatrix} x_v \\ x_m \end{bmatrix}$$



Estimation for SLAM

The estimate is:

$$\vec{x} = \begin{bmatrix} x_v \\ x_m \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} P_v & P_{vm} \\ P_{mv} & P_m \end{bmatrix}$$

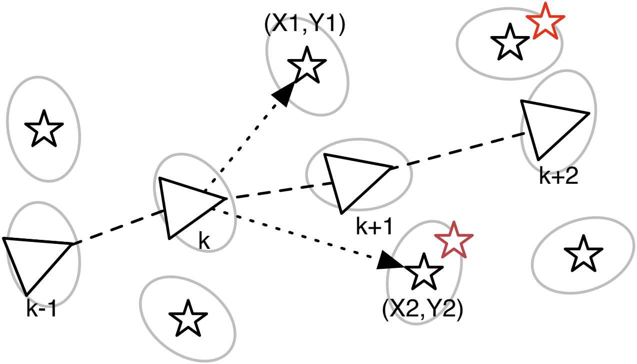
Where:

$$\mathbf{F} = \begin{bmatrix} F_v & \emptyset \\ \emptyset^T & I \end{bmatrix}$$

The complexity of updating is typically $O(N^2)$, where N is the number of map features. This can cause a problem for large maps!

Data Association

When detecting features there is a need to determine correspondence:



- Known feature
- New feature

Known features

- The map can be updated partially / sequentially. Only the vehicle estimate and the relevant feature are updated
- The matched features can be collected in a batch and processed as a partial update
- The missing features are kept but the uncertainty is not updated as this could give a “false” sense of security

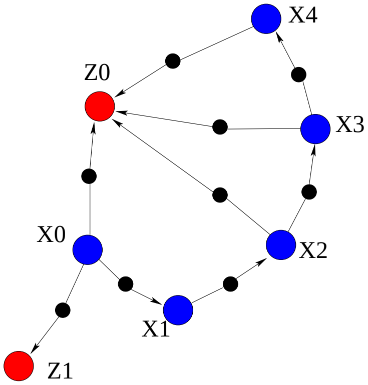
New Features

New features should *not* be accepted directly, as spurious data should be kept out of map, i.e., moving people, cars, . . .

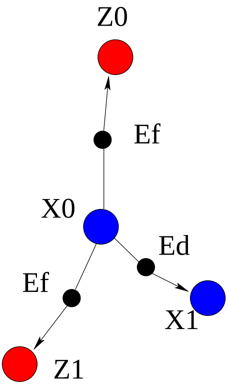
1. Keep a list of detected but not verified features (observe N times before inclusion)
   - +: Simple to implement / Good for frequent updates
   - −: For limited FOV sensors could be a problem
2. Include immediately but observe cross correlation and save a history to allow roll-back
   - +: efficient use of information / good for slow updates
   - −: complicated to implement

Robust SLAM

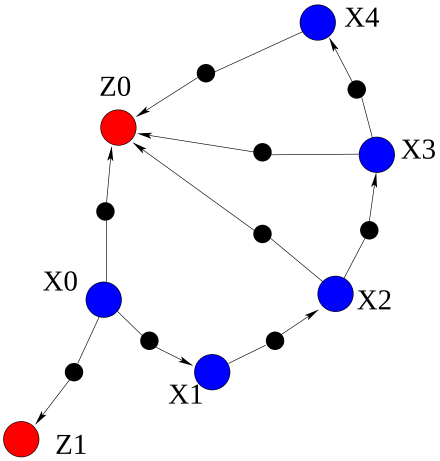
- Rephrasing EKF problem using a graphical model



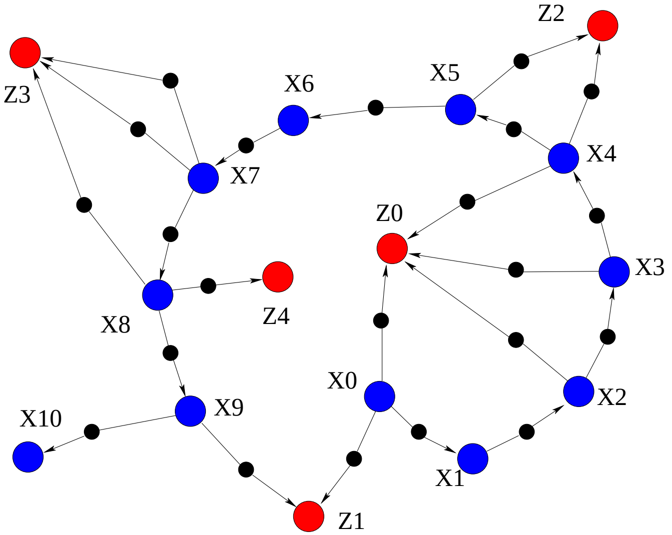
Starting a model



Entering more data



A graphical model example



Graphical SLAM Motivation

- Addressing the problem of flexible data association
- Addressing linearisation challenges
- Imposing topological constraints

Problem statement

- $\{x_i\}$ the robot path (set of poses), $(i \in \{1 \ldots N_p\})$
- $\{z_j\}$ feature coordinates $(j \in \{1..N_m\})$
- $\{d_i\}$ dead reckoning measurements, between feature measurements
- $\{f_k\}$ feature measurements, $(k \in \{1..N_f\})$
- Λ the $f \leftrightarrow z$ association

$$P(x, z, d, f, \Lambda) = P(d, f | x, z, \Lambda) P(x, z, \Lambda)$$

Probabilistic model

$$
\begin{aligned}
P(x, z, d, f, \Lambda) &\propto P(d, f|x, z, \Lambda)P(x, z, \Lambda) \\
P(d, f|x, z, \Lambda) &\propto P(d|x)P(f|x, z, \Lambda) \\
P(x, z, \Lambda) &\propto P(\lambda) = P(N_f) \propto e^{-\lambda N_f} \\
P(x, z, d, f, \Lambda) &\propto P(d|x)P(f|x, z, \Lambda)e^{-\lambda N_f}
\end{aligned}
$$

An energy model

- Definition of energy/entropy of the model:

$$E(x, z, d, f, \Lambda) = -\log(P(d|x)) - \log(P(f|x, z, \Lambda)) + \lambda N_f$$

- Or $E(x, z, d, f, \Lambda) = E_d + E_f + E_\Lambda$
- Or: ...

$$E(x, z, d, f, \Lambda) = E_d(x) + E_f(x, z) + E_\Lambda(n_j) \quad (1)$$

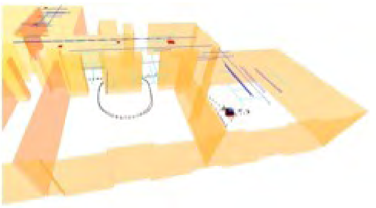
$$E_d = -\sum_{i=1}^{N_p} log(P(d_i|x_{i-1}, x_i)) = \frac{1}{2}\sum_{i=1}^{N_p} \xi_i^T k_i \xi_i \quad (2)$$

$$E_f = -\log(P(f|x, z, \Lambda)) = \frac{1}{2}\sum_{k=1}^{N_m} \eta_k^T k_k \eta_k \quad (3)$$

$$E_\Lambda = -\sum_{j=1}^{N_f} \lambda(n_j - 1) \quad (4)$$

$$\xi_i = T(x_i|x_{i-1}) - d_i \qquad \eta_k = h(T(z_j|x_i)) - f_k$$

SLAM Example

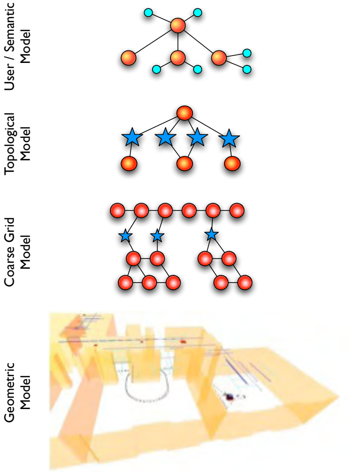


- M-Space (Folkesson et. al. 2005)
- G-SLAM (Folkesson & Christensen 2004)
- Vision, Laser Integration
- Extensively tested
- "Open"-Source Toolkit (CURE)
- (Video example)

What representations are needed?

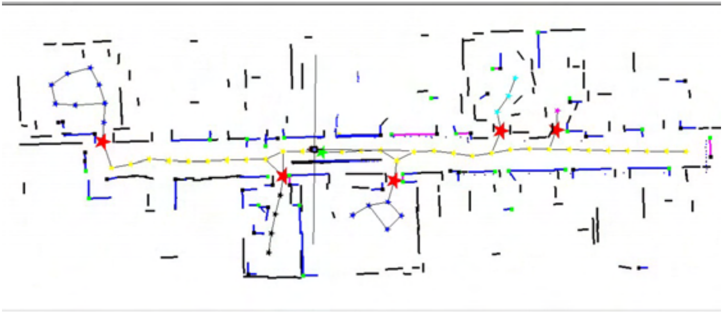
- A semantic model of “user” labels
- A topological map - “chunking” of space
  - A way to organise maps to make concept more scalable
- A coarse grid map for intermediate processing (see later)
- A metric representation to allow “real” localisation

Representations



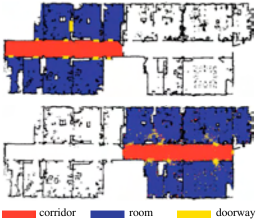
- Semantic model
  - Rooms
  - Places
- Topological
  - Doors
  - Gateways
- Coarse Grid
  - Basic grid
- Graph SLAM

Coarse Grid Map



(Video)

Room Classification



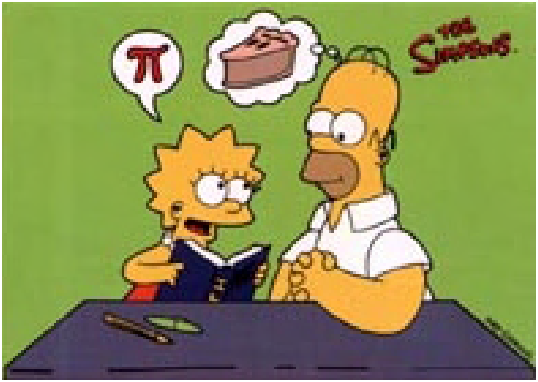
- Martinez & Burgard (2005)
- Simple classifiers
- AdaBoast for combination
- Room, Corridor, & Door
- Detects room changes
- Early detection of room category

Language Processing

- Spoken dialogue for human interaction
- Specification of locations
  - "This is the hallway"
- Query on location, status
  - "Where are we?"
- Basic Motion
  - "Follow me"
  - "Goto the kitchen"
- Correction
  - This is the corridor / not the hallway!"

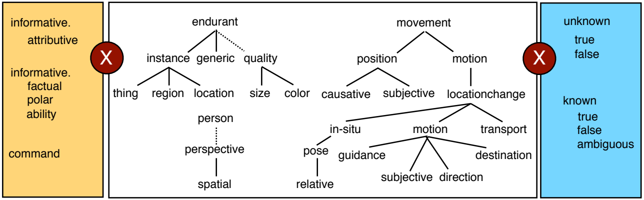
We need an ontology!

- The ontology is needed to assign semantics to spoken phrases

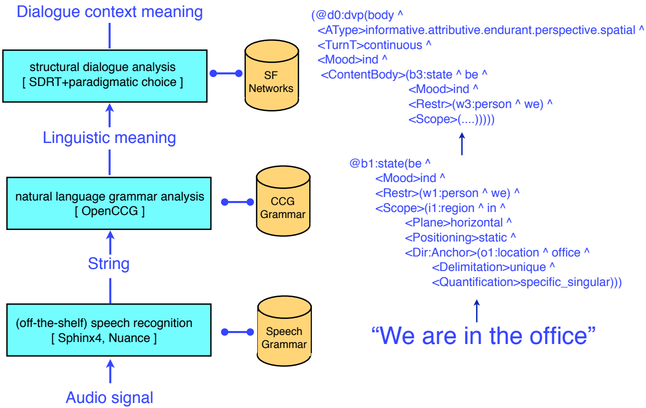


Ontology based mediation

- Ontology based cross-modal content association
  - Meaning: propositional content + intention + context update
  - Ontology for propositional truth and for intentions
  - Multi-valued truth system for update/control

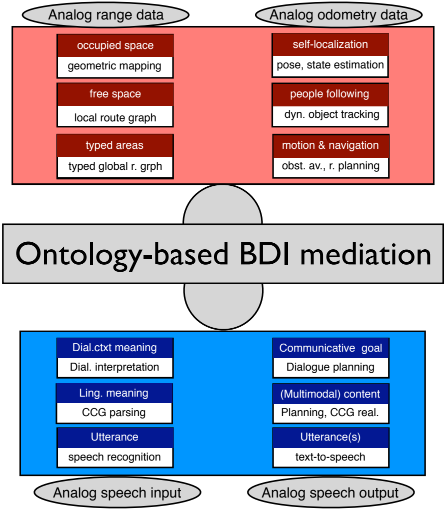


Spoken Language Processing



Explorer Architecture

CoSy @ Onassis
H. I. Christensen
Introduction
Early Evaluation
Explorer
Social Interaction
Perspectives



Early Experiments

- Evaluation – live experiments at KTH & DFKI
- Evaluation of the different phases
- Brief run through to give an impression of functionality

This is the hallway

Processing of data

SpLAM

dyn. object tracking: found dyn. object → people following: success

start head-tracking

mapping: ?unknown area, create area node, add struct.descr. → mapping: successful update

BDI

command.movement. motion.locationchange. motion.guidance.person

response to command: success

informative.attributive. endurant.instantiation (location)

response to assertion: success (unknown_t)

ComSys

command (imperative)

@f1:action(follow ^
<Mood>imp ^
<Actor>(r1:hearer ^ robot) ^
<Patient>(i1:person ^ I))

produce positive grounding feedback

assertion (indicative)

@b1:state(be ^
<Mood>ind ^
<Restr>(t1:thing ^ this ^ ...) ^
<Scope>(o1:location
^ hall ^ ...)

produce positive grounding feedback

Start guidance
"Come with me"

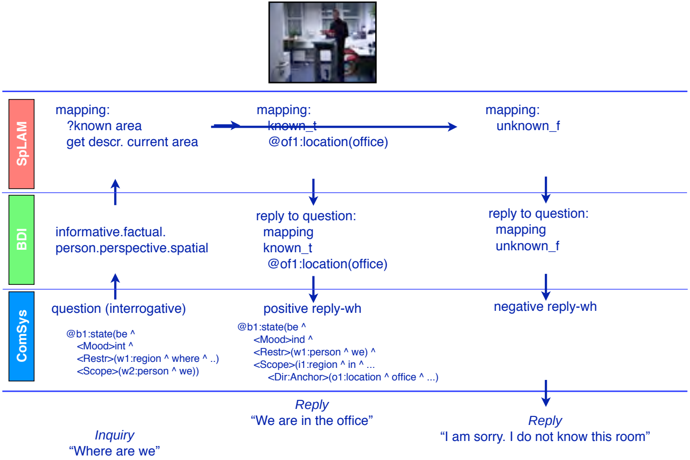
Acknowledge
"Okay." [Look]

Instruction
"This is the hallway"

Acknowledge
"Yes."

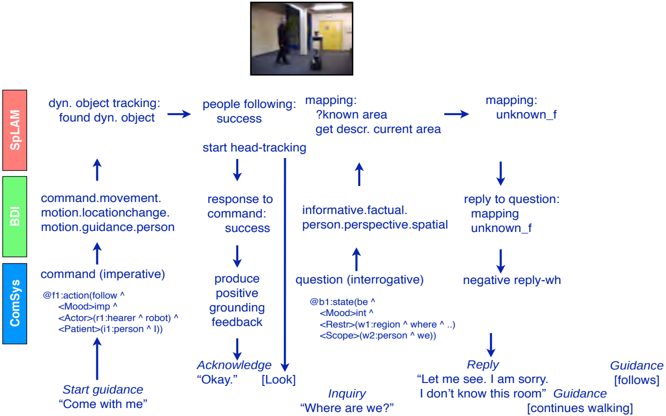
Where are you?

Processing of data



Parallel Control

Processing of data



Explorer Summary

- Early integration of dialogue behaviour with spatial mapping
- Enable user centred mapping strategy
- Integration of traditional spatial mapping and qualitative understanding
- Future work
  - Extension to 3D for spatial modelling
  - Integration of vision for recognition
  - Use of diectic gesture information

Outline

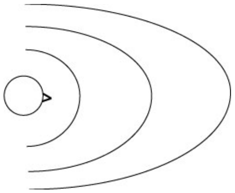
Dynamic interaction

- Interaction patterns for casual encounters in public spaces
- Hypothesis: pattern of interaction similar to person-person encounters
- Interaction patterns have been described using proxemics
- Objective
  - To design a pattern of interaction based on proxemics

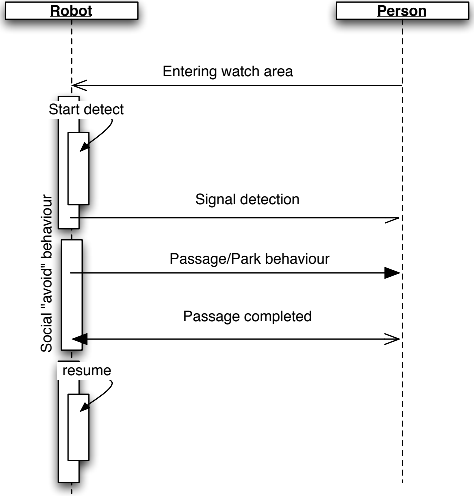
Proxemics

Introduced by Hall 1966 as a model for personal spaces

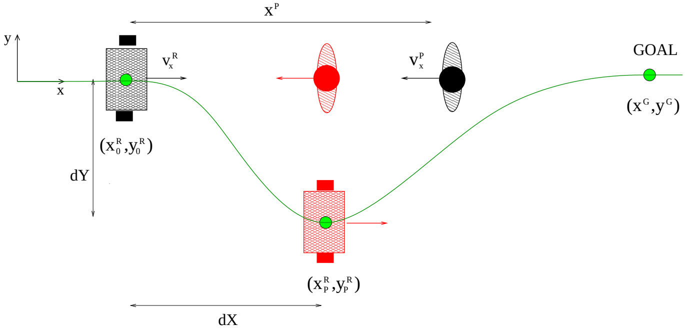
- Four areas to be considered
  - Intimate: < 45 cm
  - Private: 45–120 cm
  - Social: 120–300 cm
  - Public: > 300 cm



An interaction pattern

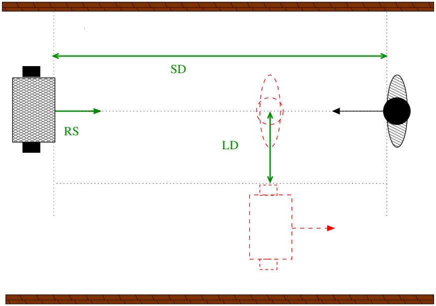


Design of the interaction strategy



Issues to be considered in the design

- When to start the avoid behaviour?
- How to signal detection?
- What is an acceptable speed of interaction?
- What is the minimum acceptable distance for passage?

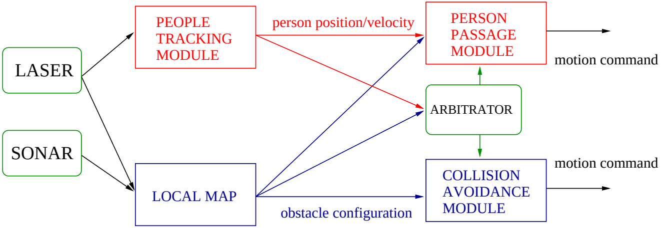


An implementation

- Tested in a hallway setting
- The interaction pattern is 1-dimensional
- Test conditions easier to control
- Implemented on a Performance PeopleBot

People detection and tracking

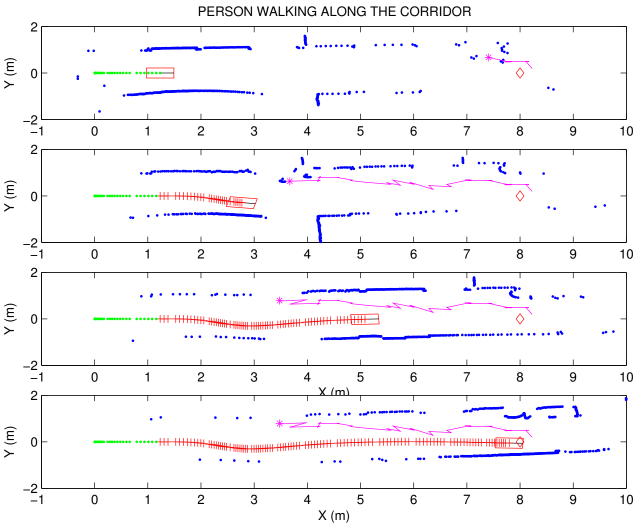
- Detection of people using a SICK laser scanner
- Particle filter based detection of people
- Relatively robust to variations person appearance over time due to continuity constraints



Example results

CoSy @ Onassis

H. I. Christensen



Evaluated in a number of user tests

- Test of passage behaviour considering
  - Distance to first detection
  - Speed of travel
  - Passage distance (in a narrow hallway)
- Early tests on 16 robot users!

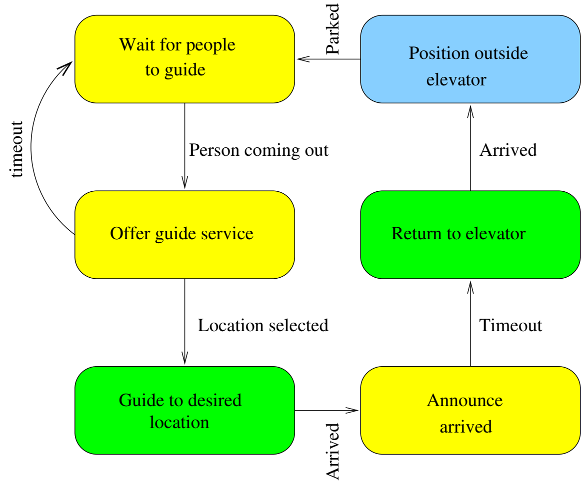
Test results

- Faster is better due to more visible interaction pattern
- Early signalling is better
- A large distance of passage is better

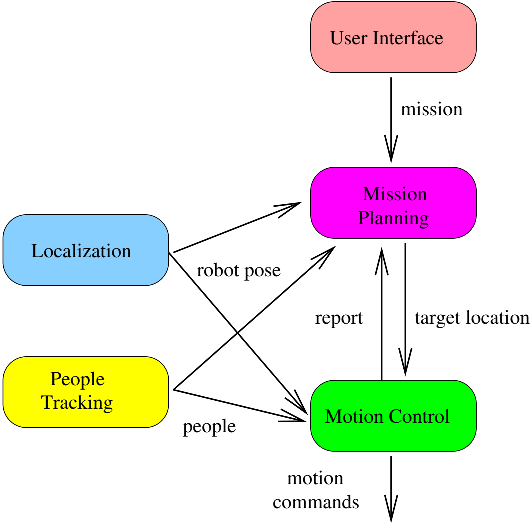
Evaluation considerations

- Evaluation of passage alone creates an artificial interaction pattern
  - You know that you are going to interact with a robot!
  - You have a sense of interaction
- A need to create a more natural setting for evaluation
- A robotic tour guide is one such domain

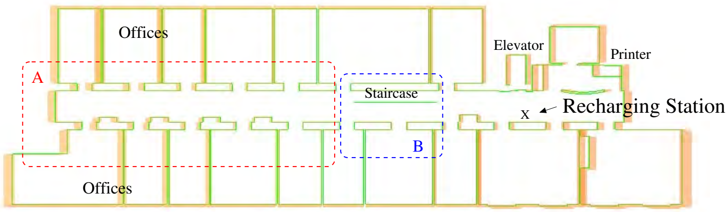
Tour Guide – State Diagram



System Components

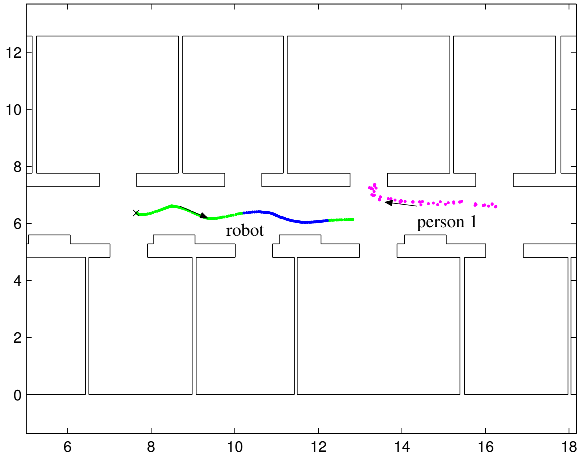


The environment

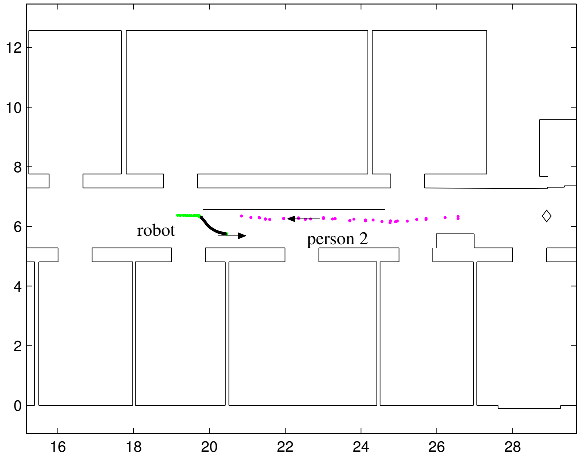


Example performance - video

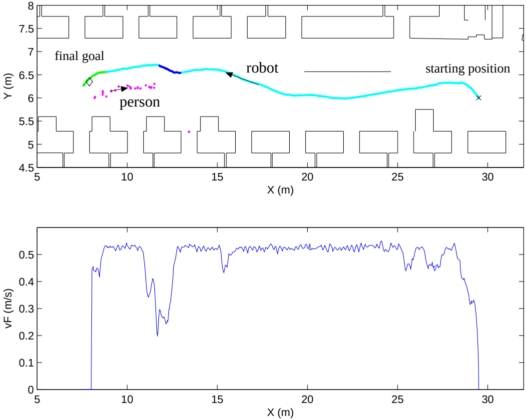
Person passage example



Park to the side example



Trajectory/Velocity Example



Trial with robot system over 2 months

Period	Missions	Pers Pass	Park	Total
1' week	28	5	32	37
2' week	38	5	29	34
3' week	26	3	25	28
6' week	33	1	34	35
8 weeks	171	21	161	182

Long-term study and main conclusions

- Questionnaire distributed to all in the lab for evaluation.
- It takes time for people to start interaction
- Most people were "comfortable" with the robot
- Tuning of parameters is a "sensitive process"
- Social rules makes environmental integration easier

Outline

Summary / Perspectives

- Design of competent systems for interaction with people is a highly interdisciplinary undertaking
  - Perception: recognition, mapping, people tracking, manipulation
  - Language: Dialog, Representations, ...
  - Reasoning: Pre-planning is not effective due to stochastics
  - User Studies: Effective interfaces
  - Systems Integration: Architectures, Representations, Evaluation

Cognitive Systems

- Made possible by advanced in computing, AI and perception
- Poses an interesting benchmark for integration
- The main challenges are
  - Making systems robust / truly autonomous
  - Making systems scalable / beyond engineering

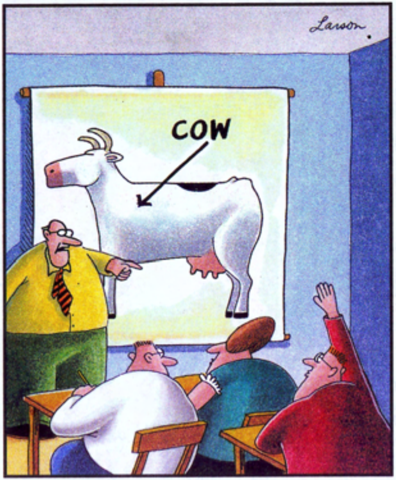
Cognitive Assistance

- The real challenge to design a system for people
  - Dialog interaction and
  - Physical interaction, say assistance
- Operating in proximity of people is demanding
  - Weight / payload ratio is a challenge
  - Many people do not have computer skills
  - Operation 24/7 within any environment

Cognitive Systems for Cognitive Assistance

- The next frontier is in the home environment
- Robustness and autonomy is crucial to success
- Vision: a robot in every home
- Price / performance is the challenge
- Especially the performance of maintaining a multi-modal dialog with people.

Questions / Comments?



"Yes ... I believe there's a question in the back."

The End