

# RoSta

## A Brief View Over Benchmarking Activities in Service Robotics

- Introduction to RoSta in general
- Introduction of WP 4 – Benchmarking
- State of the art – Benchmarking activities
- Analysis and interpretation
- Outlook and next steps

### Vision

- Initiative on the definition of formal standards and the establishment of “de facto” standards in the field of service robotics.
- Formulation of standards (action plans) in a few, selected key topics which have the highest possible impact.
- Form the root of a whole chain of standard defining activities going far beyond the specific activities of RoSta.

### 4 Topics (“Action Lines”)

- Glossary/ontology for mobile manipulation, service robots
- Specification of a reference architecture
- Specification of a middleware
- Formulation of benchmarks

<http://www.robot-standards.org>

<http://wiki.robot-standards.org>

## Objectives

- Specify/suggest a portfolio of benchmarks in robotics and formulate their implementation strategy:
  - Test cases and metrics for the assessment of specific methods, processes or robot related functions (synthetic benchmarks).
  - Application benchmarks which measure the fulfillment of challenging tasks as subsets of complete application scenarios of robot assistants or companions (task domain).
  - Competitions which aim at measuring a robot's relative performance among peers for given test cases.

„Car-Like“ approach of increasing complexity and abstraction:

- Easy to measure and understand basic values:

HP, fuel consumption, engine size, weight, CO2 exhaust, ...

- More complex and abstract values, still understandable:

Slalom behaviour, elk test, Minimum break distance, trip time for northern loop of Nürnburgring, ...

- Abstract and subject evaluation, understandable but discussable:

Design, drive feeling, over all comfort, family suitable, ...

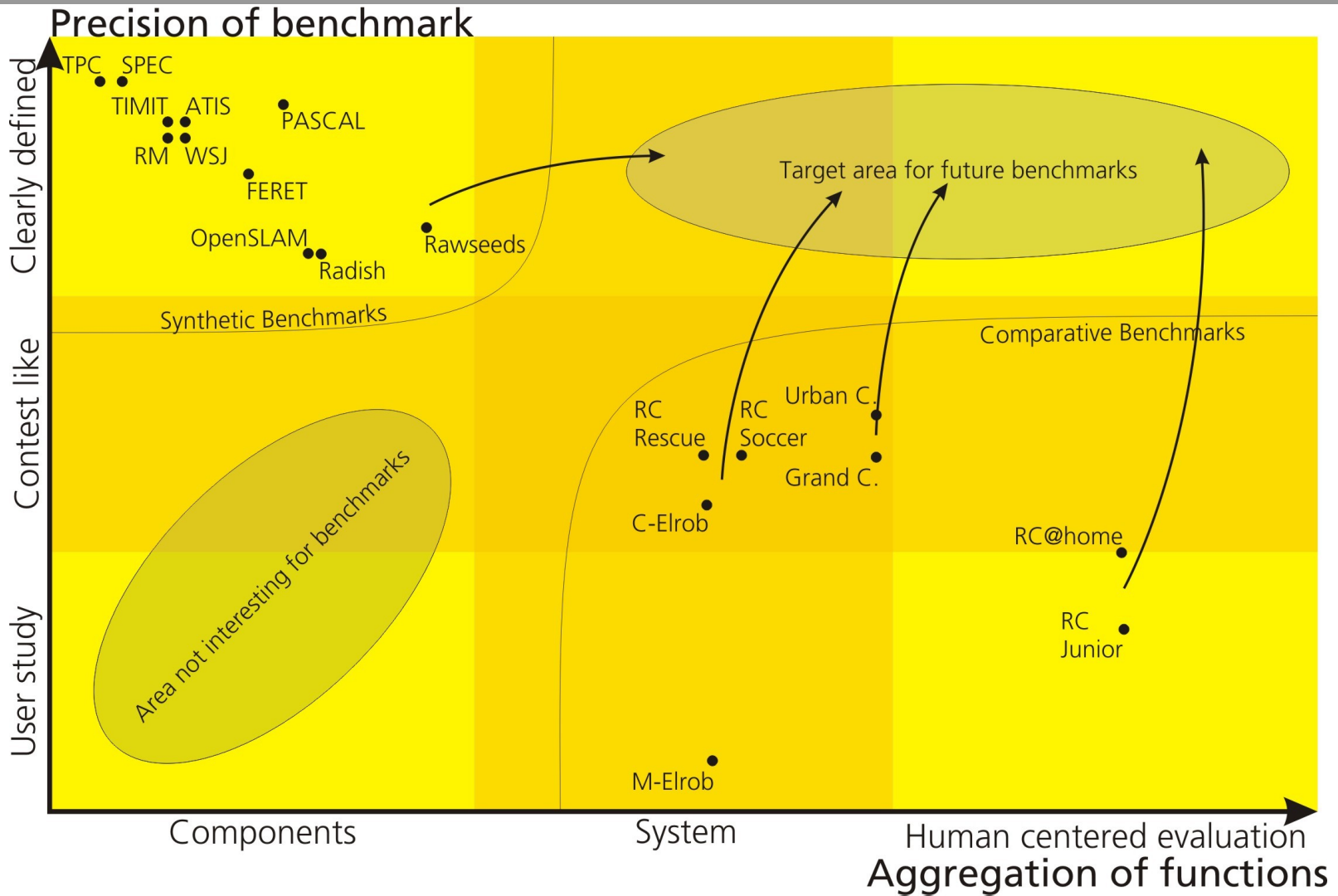
## Three tasks

- Task 4.1 Compilation and Evaluation of State of the Art in benchmarking service robots
- Task 4.2 Requirement analysis
- Task 4.3 Action plan and recommendation

- Research and expert meetings identified:
  - 8 service robot related competitions
  - 11 benchmarks of service robot related technologies
    - Navigation
    - Object recognition
    - Speech recognition
    - Computer industry and science
    - Manipulation and grasping

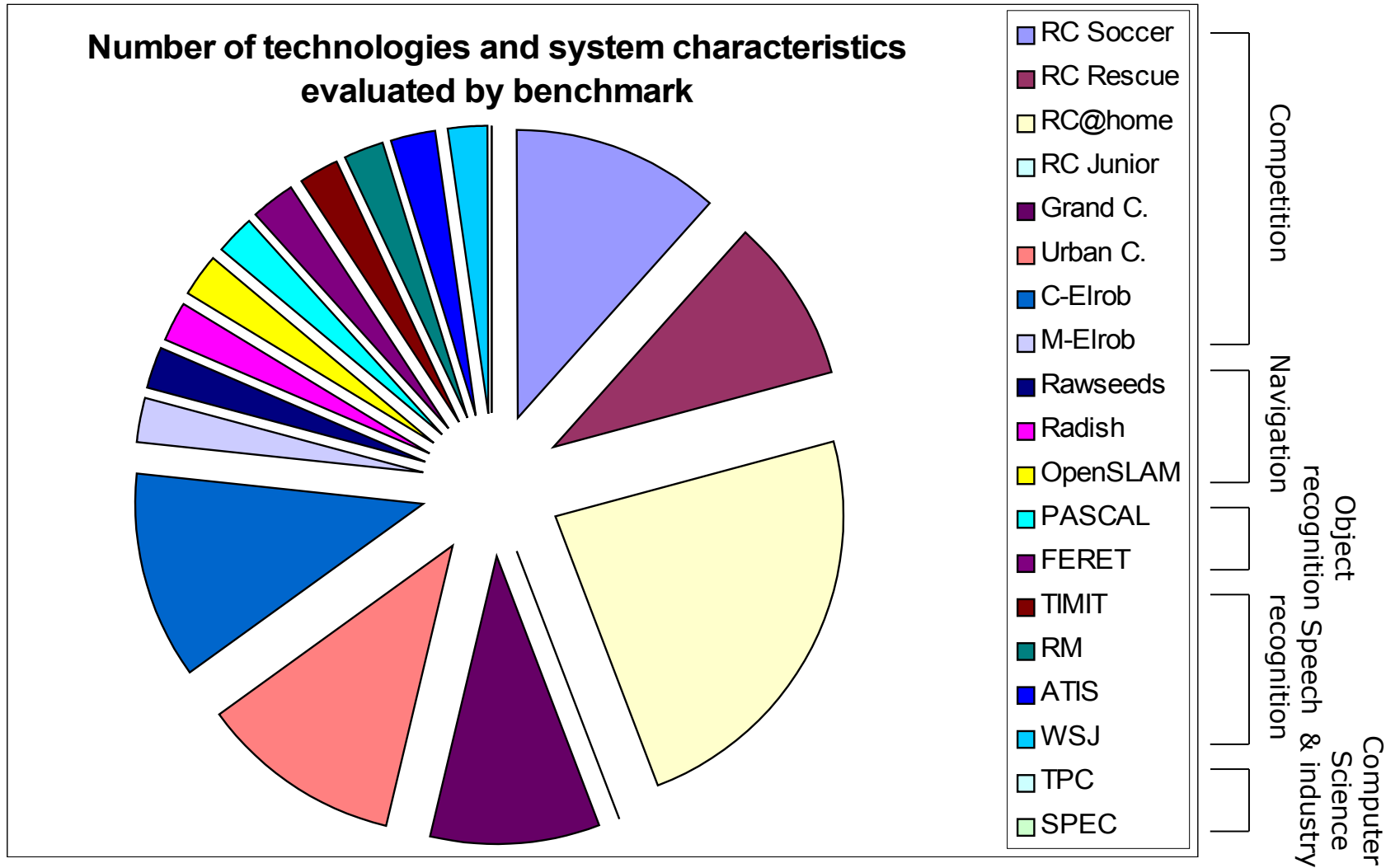


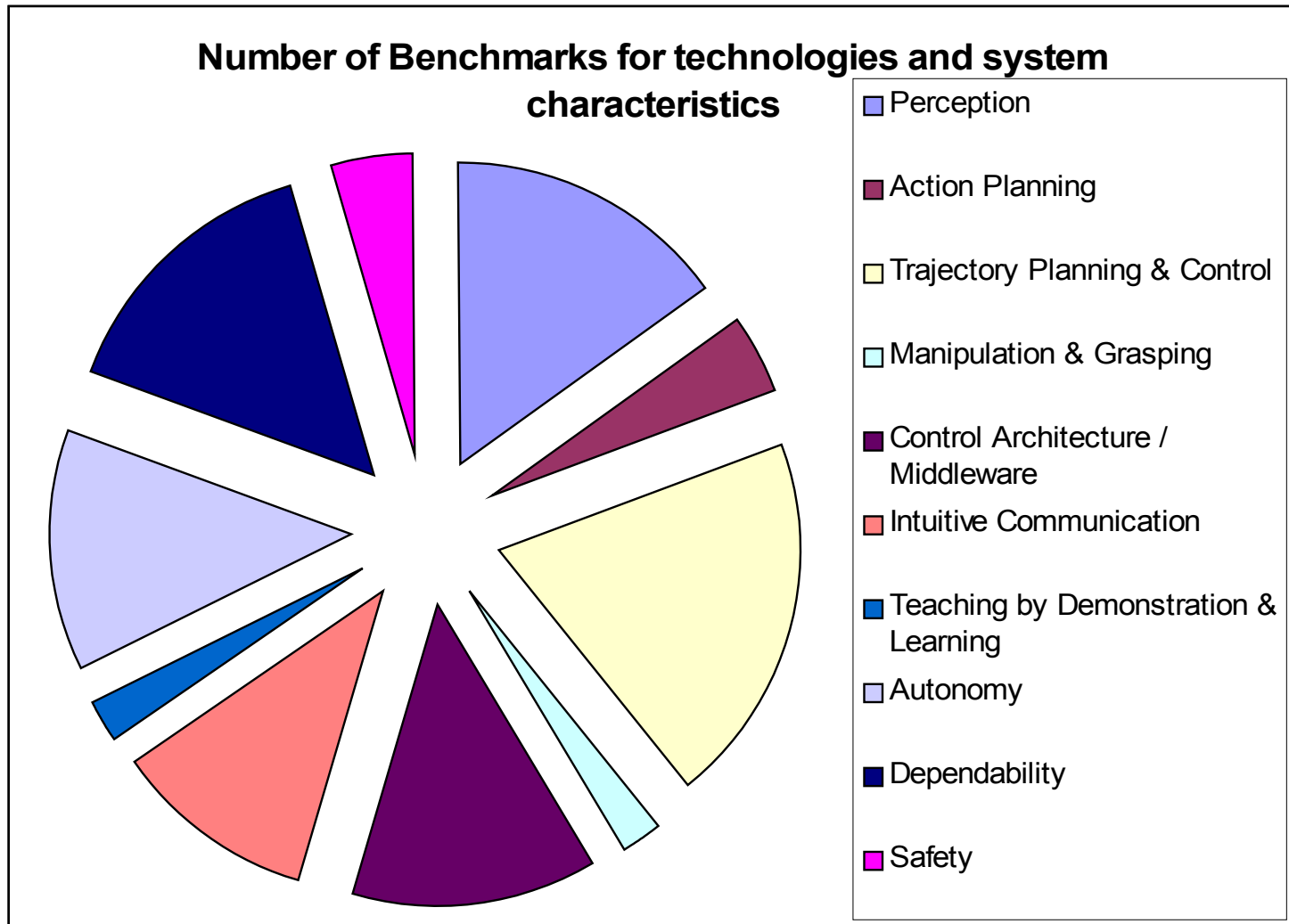
- Evaluating the benchmarks
- Definition of a metric for identified benchmarks
  - Precision
    - User study
    - Contest like
    - Clearly defined
  - Complexity
    - Components
    - System
    - Human centered evaluation
- Allocation of benchmarks within metric

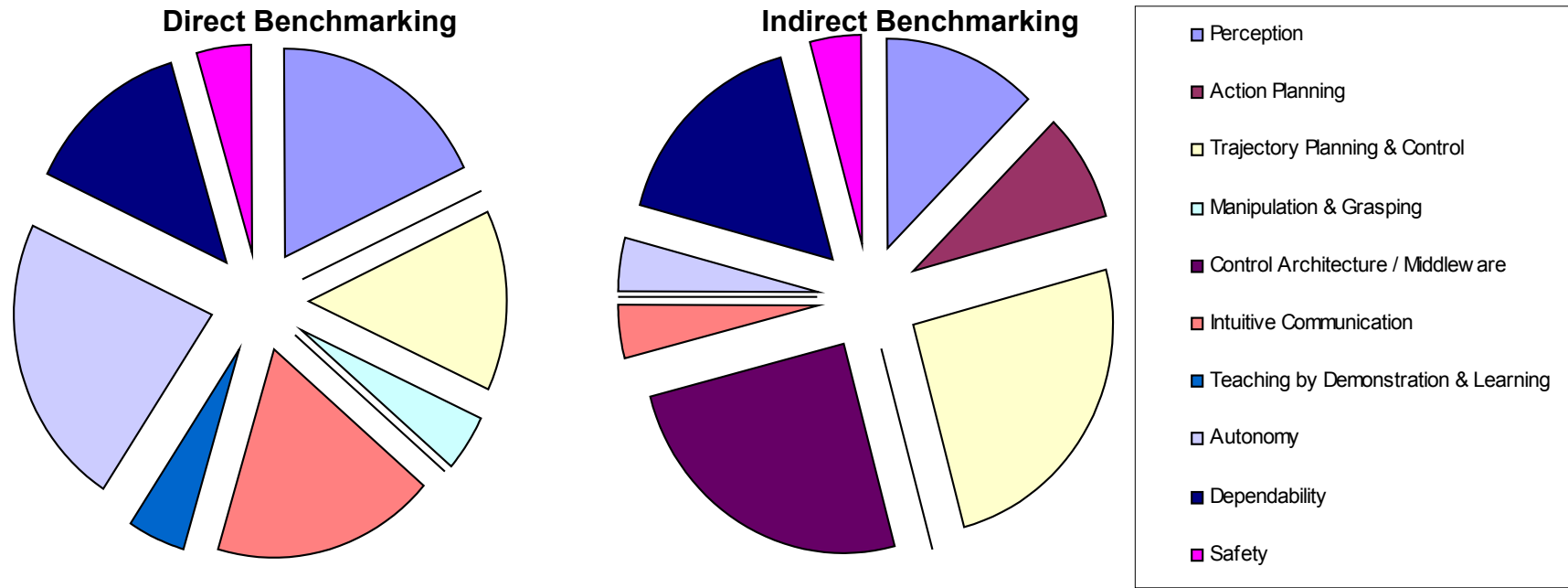


- Key technologies and system characteristics:
  - Perception
  - Action Planning
  - Trajectory Planning & Control
  - Manipulation & Grasping
  - Control Architecture / Middleware
  - Intuitive Communication
  - Teaching by Demonstration & Learning
  - Autonomy
  - Dependability
  - Safety

- Benchmarks examined by evaluated technology
  - Number of technologies and system characteristics evaluated by identified benchmarks
  - Number of benchmarks evaluating technologies and system characteristics
  - Comparison between direct and indirect benchmarks
    - Direct: Technology or system characteristic is directly addressed by benchmark
    - Indirect: Technology or system characteristic influences benchmark result







## Interpretation of results

- There is a lack of benchmarks in Action Planning, Manipulation & Grasping, Teaching by Demonstration & Learning, and Safety
- Only competitions address more than one key technology
- System characteristics are only addressed by competitions
- There are no direct evaluations for Action Planning and Control Architecture / Middleware

Remark: The quality of the benchmarks and competitions was not taken into account!



- Empty spaces within the metric have to be filled
- Every component which is subject to research must be evaluable with a commonly used benchmark
- Specify evaluation plans for components both for single evaluation and for evaluation in combination with other components
  
- Questionnaire for requirement analysis
- Collect results and evaluate / discuss with expert group
- Depending on feedback individual interviews with key players in robotics (scientific and industry)

<http://www.robot-standards.org>

<http://wiki.robot-standards.org>

## Questionnaire to identify requirements for benchmarking

- Which technologies (navigation, manipulation, MMI ...) of mobile service robots should be measured with a standardized benchmark?
- What benefits/drawbacks of establishing standardized benchmarks for mobile service robots can you think of?
- How would you benchmark "robustness" and "dependability" for a service robot?
- How would you benchmark "autonomy" for a service robot?
- How would you split a mobile service robot concept? Please name possible sub-divisions:
- What is needed (central benchmarking center, simulators, standardized open architectures ...) to generate a living benchmarking culture?
- Which (European and international) stakeholders should be involved in a public consolidation workshop "Benchmarks in Robotics"?