

Refinement of test procedures and evaluation criteria

This document refines the evaluation methods for the competition that were introduced in the technical annex of the call (<http://evaal.aalooa.org/current-competition/techannex>). Specifically, it refers to the procedures and parameters of the tests and to the criteria "Accuracy", "Installation complexity", and "Availability".

Remember that each localization system will be evaluated in two phases:

“

Phase 1. In this phase each team must locate a person inside an Area of Interest (AoI). The AoI in a typically AAL scenario could be inside a specific room (bathroom, bedroom), in front of a kitchen etc. The AoI will be disclosed to the competitors before the competition, as early as possible.

Phase 2. In this phase a person that moves inside the Living Lab must be located and tracked (we plan only 2D localization and tracking here). During this phase only the person to be localized will be inside the Living Lab.

In this phase each localization system should produce localization data with a frequency of 1 new item of data every half a second (this will be also used to evaluate availability). The path followed by the person will be the same for each test, and it will not be disclosed to competitors before the application of the benchmarks.

“

1. Organization and Test Procedures

A few notes about the organization of the competition:

- The organizers give an appointment to each competitor so that each competitor may stay at the living lab for the time necessary for the evaluation of his/her system. The schedule is available in Section 3 of this document.
- Only one competitor at a time is admitted in the living lab (no other competitors will be admitted during the evaluation of the other competitors). The reason is that we will use the same paths for the tracking tests of all competitors, and such paths cannot be disclosed before the actual test because this might invalidate the results.
- The path to be tracked will not be disclosed to anybody before entering the living lab for the competition
- The areas of interest are described in the Annex 1.
- The criteria "Availability" and "Accuracy" will be performed automatically with the supervision of the evaluation committee. The evaluation committee will evaluate the other criteria: "Installation Complexity", "User acceptance" and "Integrability in AAL systems".
- We will give the opportunity to each competitor to repeat the test one time and to opt

- for the best result between the first try and the second.
- The organizers will not provide any tools for bricolage, hence competitors should bring all the tools they need. If necessary, there is a mall 10 minutes by car from the living lab where the competitors may buy some tools. **In any case, if competitors need to hang devices on the walls and they plan to use glue, nails or any other method that may damage the living lab they are requested to contact the organizers as soon as possible.**

Installation of localization systems at the living lab:

- Each competitor will be given up to 3 hours to: install the system, execute the tests, uninstall everything. **Specifically, we plan to give at most 1 hour for the installation. If any competitor feels that he/she cannot meet this requirement, he/she is invited to communicate this as soon as possible to the organizers.**
- After the installation the organizers will make a preliminary test to ensure that the connection between the competing system and the evaluation tools is correctly established. The time required to solve any issue related to this does NOT count for the Installation Complexity criteria.

In all the tests the competing systems will localize the movements of an actor (a person of the organization trained in moving along pre-defined paths).

The tests in the two phases are organized as follows:

Phase 1:

Each system is requested to identify 5 Areas of Interest (AoI). The actor will move along random paths and will stop in each AoI for 30 seconds.

For details about the AoI please refer to Annex 1.

Phase 2:

Each system is requested to track the actor along three different paths (the paths are the same for all competitors). The evaluation criteria Accuracy and Availability will be computed on the three paths aggregated. Each path will last up to a couple of minutes. To help the competitors we give a sample path (that will not be used during the competition) in the Annex 1.

2. Refinement of Criteria

Accuracy:

Each localization system will produce a stream of tuples: <time, x, y, area_of_interest_code>, one sample every half a second.

Phase 1) The user will stop (after a random walk) 30 seconds in each Area of Interest (AoI). Accuracy in this case will be measured as the fraction T of time in which the localization system provides the correct information about:

- Presence of the user in a given AoI
- Absence of the user from any AoI

The score is given by:

$$\text{Accuracy score} = 10 * T$$

Phase 2) The stream <time,x,y> produced by competing systems will be compared against a

logfile of the expected position of the user. Specifically, we will evaluate the individual error of each measure (the Euclidian distance between the measured and the expected points), and we will estimate 75th percentile P of the errors. In order to produce the score, P will be scaled in the range [0,10] according to the following formula:

Accuracy score = 0	if	$P \geq 3 \text{ m}$
Accuracy score = 10	if	$P \leq 0,5 \text{ m}$
Accuracy score = $4 \cdot (0.5 - P) + 10$	if	$0,5 \text{ m} < P \leq 2 \text{ m}$
Accuracy score = $2 \cdot (4 - P)$	if	$2 \text{ m} < P \leq 4 \text{ m}$

The final score on accuracy will be the average between the scores obtained in phase 1 and in phase 2.

Installation Complexity:

Thus measures the time T necessary to install the localization system. The time T is measured in minutes from the time in which the competitor enter in the living lab to the time when they declare they completed the installation (no further operations/configurations of the system will be admitted after that time), and it will be multiplied by the number of people N working on the installation.

The parameter T*N will be translated in a score (ranging from 0 to 10) according with the following formula:

Installation Complexity Score = 10	if $T \cdot N \leq 10$
Installation Complexity Score = $10 \cdot (60 - T \cdot N) / 50$	if $10 < T \cdot N \leq 60$
Installation Complexity Score = 0	if $T \cdot N > 60$

Availability:

Availability A is measured as the ratio between the available samplings produced by the localization system and the expected samples.

In both the first and second phase, each localization system is expected to provide 1 sample every half a second, hence the number of expected samplings is given by the duration of the test * 2.

The values of availability A will be translated into a score (ranging from 0 to 10) according to the following formula:

$$\text{Availability score} = 10 \cdot A$$

3. Schedule of Competitors

July 26th 2011

15:00 – 18:00 CapFloor (Andreas Braun)

July 27th 2011

8:00 – 11:00	SNTUmicro (Igor Shirokov, Sevastopol National Technical University)
11:00 – 14:00	AIT (Thomas Fuxreiter, AIT)
15:00 – 18:00	GEDES-UGR (Tomás Ruiz-López, University of Granada)

July 28th 2011

8:00 – 11:00	iLoc (Stefan Knauth, Lucerne University of Applied Sciences, iHomeLab and Stuttgart University of Applied Sciences)
11:00 – 14:00	n-Core Polaris (Dante I. Tapia, University of Salamanca)
15:00 – 18:00	OwlPS (Matteo Cypriani, University of Franche-Comte)

July 29th 2011

8:00 – 11:00	The OFFIS AAL Team (Enno-Edzard Steen, OFFIS - Institute for Information Technology)
11:00 – 14:00	Ubisense (Jaouhar Jemai, Ubisense AG)