# Evaluating AAL Systems Through Competitive Benchmarking (EvAAL) (Technical aspects of the first competition)

Stefano Chessa[1,2], Francesco Furfari[1], Francesco Potortì[1], Paolo Barsocchi[1]
Mohammad-Reza Tazari[3], Reiner Wichert[3]

[1] Istituto di Scienza e Tecnologie dell'Informazione,
Consiglio Nazionale delle Ricerche
Via Moruzzi 1, Pisa, Italy

[2] Department of Computer Science, University of Pisa
Largo Pontecorvo 3, Pisa, Italy

[3] Fraunhofer Ambient Assisted Living Alliance,
c/o Fraunhofer-Institut für Graphische

***Abstract***

EvAAL (Evaluating AAL Systems Through Competitive Benchmarking) is an international competition aimed at the evaluation and assessment of Ambient Assisted Living systems components, services and platforms. In 2011 took place the first edition of EvAAL on the special theme of Indoor Localization and Tracking for AAL. This paper describes the technical aspects of the first edition of EvAAL and draws a roadmap for the future editions.

## *1. Introduction*

The evaluation and comparison of complex Ambient Assisted Living (AAL) systems is still far from being a reality [1]. On the other hand, the evaluation and assessment of components, services, and platforms for AAL systems is essential to ensure the progress, and, ultimately, the success of AAL technologies.

EvAAL is an international competition on AAL supported by the AALOA association [2] and organized by the universAAL project [3]. It aims at advancing the state of the art in the evaluation and comparison of AAL platforms and architectures. In particular, EvAAL aims at generating an environment in which researchers, students, practitioners and industries can compare their solutions and build together methodologies and approaches that make such a comparison possible. Since at present the complexity of AAL systems makes not possible their full comparisons, EvAAL adopts a gradual approach, by dividing the problem into sub-problems, and by deferring the whole problem when the knowledge on AAL systems evaluation is more developed. Specifically, the first editions of EvAAL promote competitions on specific AAL components, in order to create data sets, benchmarks and evaluation methodologies. Then, based on the knowledge built in this phase, the subsequent EvAAL editions will focus on more complex (and possibly complete) AAL solutions.

In the first edition it was chosen to organize a single track of competition on the topic "Indoor Localization and Tracking". Localization was chosen because it is a key component of many AAL services. Recent years have witnessed an increasing trend of location-based services and applications. In most cases, however, location information is limited by the accessibility to Global Navigation Satellite Systems (GNSS), largely unavailable for indoor environments. The scope of this competition is to award the best indoor localization system from the point of view of Ambient Assisted Living (AAL) applications.

For organization reasons, EvAAL 2011 was organized in two major events: the actual competition organized at the CIAmI Living Lab in Valencia (SP) [6], on the 27th-29th July, and the concluding workshop held in Lecce on the 26th of September (the workshop was a side event of the AAL Forum [4]). This gave the opportunity to each competitor to dispose of the living lab for a long time slot (3 hours), during which install, test and uninstall his/her system.

This paper presents the technical aspects of this first EvAAL edition by discussing the evaluation criteria, the benchmarks and the results of the competition.

## *2. Evaluation criteria*

In order to evaluate the competing localization systems, EvAAL used a set of criteria weighted according to its relevance and importance for AAL applications. For each criterion, each competing artifact receives a score, that can be either measured by direct observation, or, when a direct measurement is not possible, it is determined by the Evaluation Committee, a committee composed of volunteer members of the Technical Program Committee TPC, which were present during the competition at the Living Lab.

The criteria (along with the respective weights) are the following:

*Accuracy (weight: 25%):* each produced localization sample has been compared with the reference position and the error distance has been computed. Each localization system produced a stream of tuples, one sample every half a second. Specifically, the accuracy has been evaluated for each phase as:

- Phase 1: The accuracy in this case was measured as the fraction T of time in which the localization system provides the correct information about presence or not in a given AoI, the final score was given by 10*T.

- Phase 2: The stream produced by competing systems has been compared against a logfile of the expected position of the actor. Specifically, we evaluated the individual error of each measure (the Euclidian distance between the measured and the expected points), and we estimated 75th percentile P of the errors. In order to produce the score, P has been scaled in the range [0,10] according to the following formula:

| | |
|---|---|
| Accuracy score = 0 | if P >4 m |
| Accuracy score =10 | if P <= 0,5 m |
| Accuracy score =4*(0.5-P)+10 | if 0,5m < P <= 2 |
| Accuracy score =2*(4-P) | if 2m < P <= 4 |

*Installation complexity (20%)*: a measure of the effort required to install the AAL localization system in a flat, measured by the evaluation committee as the total number of man-minutes of work needed to complete the installation. Thus measures the time T necessary to install the localization system. The time T was 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 was multiplied by the number of people N working on the installation. The parameter T*N was translated in a score (ranging from 0 to 10) according with the following formula:

| | |
|---|---|
| Installation Complexity Score = 10 | if T*N <=10 |
| Installation Complexity Score = 10 * (60-T*N) / 50 | if 10 < T*N <= 60 |
| Installation Complexity Score = 0 | if T*N >60 |

*User acceptance (20%):* expresses how much the localization system is invasive in the user's daily life and thereby the impact perceived by the user. This criteria is qualitative and was evaluated by the evaluation committee taking into account a predefined list of questions.

*Availability (15%):* fraction of time the localization system was active and responsive. It is measured as the ratio between the number of produced localization data and the number of expected data. In both, first and second phases, each localization system was expected to provide one 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 has been translated into a score (ranging from 0 to 10) according to the following formula:

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

*Integrability into AAL systems (10%):* The score ranging from 0 to 10 was given by the EC: 2 points for availability of libraries for integration; 2 points for use of open solutions for libraries; 2 points for use of standards; 2 points for availability of tools for testing/monitoring the system; 1 point for availability of sample applications; 1 point for availability of documentation.

### *3. Benchmarks*

The score for measurable criteria for each competing artefact has been evaluated by means of benchmark tests. To this purpose each competing team has been allocated a time slot of three hours, during which the benchmark tests had been carried out. The benchmark consists of a set of tests, each of which contributes to assessment of the scores for the artefact. The Evaluation Committee controlled all the operations to ensure a fair evaluation of each artefact.

The time slot assigned to each competitor was divided in three parts:

- In the first part, the competing team deployed and configured their artefact in the living lab. This part should last no more than 60 minutes and its duration is measured in order to produce the score for installation complexity criteria.
- In the second part, the benchmark is applied. During this phase the competitors had the opportunity to perform only short reconfigurations of their systems. In any case, this part should be concluded in 60 minutes.

- In the last part, the competitors must remove the artefact from the living lab in order to enable the installation of the next competing artefact.

Competing teams who failed to meet the deadlines in part 1 have been given the minimum score for the installation complexity criteria.

During the second part, the localization systems had been evaluated in two phases:

**Phase 1**. In this phase each team must locate the user (impersonated by an actor) 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. Each system is requested to identify 5 Areas of Interest (AoI) (see Figure 1). The actor moved along random paths and stopped in each AoI for 30 seconds.

**Phase 2**. In this phase the artefacts should localize and track the actor that freely moves in the Living Lab. During this phase only the actor to be localized was inside the Living Lab. In this phase each localization system produced localization data with a frequency of one new item of data every half a second (this has also been used to evaluate availability). Each

system was requested to track the actor along three different paths (Figure 2) that was the same for each test, and it was not disclosed to competitors before the application of the benchmarks. The first path was 36 steps length, the second path 52, and the last one 48. Moreover, all the paths were characterized by 3 waiting points, i.e. the actor stayed in the same position for 10 seconds. The evaluation criteria Accuracy and Availability have been computed on the three paths aggregated. Each test lasts up to a couple of minutes.


**Figure 1.** The Areas of Interest deployed in the Living Lab

**Figure 2.** The three different paths:
path 1 (orange line), path 2 (green line), and path 3 (magenta line)

### *4. Results*

At the Ciami Living lab 6 teams challenged themselves at the competition, namely n-core Polaris (from the University of Salamanca), AIT (from Austrian Institute of Technology), iLoc (from Stuttgart University of Applied Sciences and iHomeLab at Lucerne University of Applied Sciences), OwlPS (from University of Franche-Comte), GEDES-UGR (from University of Granada), and SNTUmicro (from Sevastopol National Technical University). Table 1 summarizes the scores got by the different competitors. In particular, the n-Core reached a best overall scores, since it received the best score for availability, installation complexity and user acceptance. Since this localization system is based on Received Signal Strength (RSS)

the accuracy score was third with respect to the other systems. The best localization system with respect to the accuracy score was AIT with the infrared technology, followed by the ultrasound devices of iLoc. The n-Core team won since it was the system that, on average, obtained an high score in all the metrics, while AIT and iLoc obtained low scores for availability and installation complexity, respectively.

**Table 1**. The final scores of competing artefacts

| *Competitor* | *Accuracy* | *Availability* | *Installation Complexity* | *User Acceptance* | *Integrability in AAL* | *Final score* |
|---|---|---|---|---|---|---|
| **n-Core** | 5,9611 | 9,8756 | 10 | 7.625 | 6.5 | **7.14** |
| **AIT** | 8,4540 | 1,3674 | 6,82 | 6,875 | 8,5 | **5,90** |
| **iLoc** | 7,8007 | 9,3922 | 0 | 5,875 | 4,5 | **4,98** |
| **OwIPS** | 1,3653 | 9,4337 | 8,4733 | 6,5 | 1 | **4,85** |
| **GEDES-UGR** | 1,8055 | 9,0193 | 0 | 6 | 10 | **4,00** |
| **SNTUmicro** | 0 | 0 | 10 | 4,375 | 3 | **3,17** |

## *4. Conclusions*

The first edition of EvAAL involved the participation of a good number of teams, and provided many feedbacks to the organizers for the next editions. We are now planning EvAAL 2012, which will open to new tracks (while keeping indoor localization). In order to improve EvAAL we have prepared and distributed a call for ideas aimed at researcher, technician, or even user. The purpose of the call for ideas is to collect suggestions for the improvement of the technical and organization aspects of EvAAL, and to collect proposals for new topics. The call for ideas can be downloaded at the EvAAL website [5]. We conclude with our warm invite inviting to everybody to help us make EvAAL a stable and widely recognised event for AAL.

## *References*


[1] CONNELLY, K., SIEK, K.A., MULDER, I., NEELY, S., STEVENSON, G., KRAY, C., Evaluating Pervasive and Ubiquitous Systems, IEEE Pervasive Computing 7(3):85-88, 2008.

[2] AALOA association, http://www.aaloa.org

[3] EU FP7 universAAL project, http://www.universAAL.org

[4] AAL Forum, http://www.aalforum.eu

[5] EvAAL website: http://evaal.aaloa.org

[6] CIAMI Living Lab website: http://www.ciami.es/valencia/