Comparing AAL Indoor Localization Systems

Paolo Barsocchi[1,⋆], Francesco Potortì[1],
Francesco Furfari[1], and Alejandro M. Medrano Gil[2]

[1] ISTI Institute of CNR, Pisa Research Area, via Moruzzi 1, I-56124, Pisa, Italy
{paolo.barsocchi,potorti,furfari}@isti.cnr.it
[2] Life Supporting Technologies, Universidad Politècnica de Madrid, Avenida Complutense n. 30, 28040, Madrid, Spain
amedrano@lst.tfo.upm.es

Abstract. Evaluating Ambient Assisted Living (AAL) systems is challenging due to the complexity and variety of solutions adopted and services offered. EvAAL is an international competition aimed to address this problem by evaluating and assessing the AAL 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.

Keywords: AAL, localization, tracking.

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. EvAAL aims at contributing to AAL disciplines in the same way as other competitions have contributed to their respective areas. Under this respect EvAAL is inspired by successful competitions such as the Trading Agent Competition [4] (TAC) and DARPA Grand Challenge [5]. In contrast with the above mentioned competitions, and beyond supporting the growth of the AAL community, the main technical objectives of the competitions organized by EvAAL are to i) enable the comparison of different AAL solutions, ii) experiment with benchmarking and evaluation methods, iii) identify relevant AAL problems, requirements and issues, and iv) identify new and original solutions for AAL. EvAAL aims at enabling the comparison of different AAL solutions, by establishing suitable

⋆ This work was supported in part by the European Commission in the framework of the FP7 project universAAL under Contract 247950.

S. Chessa and S. Knauth (Eds.): EvAAL 2011, CCIS 309, pp. 1–13, 2012.

benchmarks and evaluation metrics that will be progressively refined and improved with time. In particular, EvAAL will focus not only on comparison of algorithms, but also of cost, deployment effort, user acceptance, and soon. 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 split into 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 [7]). 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. In particular, Section 2 describes the benchmark tests we created to evaluate each competing localization system, section 3 shows the evaluation criteria used. Section 4 describes the localization system chosen as reference. The collected amount of data that are useful as benchmarks to the researcher communities are explained in section 5. While the protocol we followed during the competition is explained in section 6. The final results as well as the lesson learned in the first edition of EvAAL, and the conclusions are described at the end of this work.

2 Benchmarks

The score for measurable criteria (described later in section 3) for each competing artefact was evaluated by means of benchmark tests. For 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. For the evaluation, an Evaluation Committee (EC) was set up, composed of volunteer members of the Technical Program Committee. This EC was present during the competition and 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 (see section 3).
– 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 the first part 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 had to 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 was requested to identify 5 Areas of Interest (AoI) (see Figure 1). The actor moved along predefined paths and stopped in each AoI for 30 seconds.

Fig. 1. The Areas of Interest deployed in the Living Lab

– Phase 2. In this phase the artefacts had to localize and track the actor that freely moved in the Living Lab. During this phase only the actor to be localized was inside the Living Lab. Each localization system produced localization data with a frequency of one new item of data every half a second. Each system was requested to track the actor along three different paths (Figure 2) which were 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. Each test lasted up to a couple of minutes.

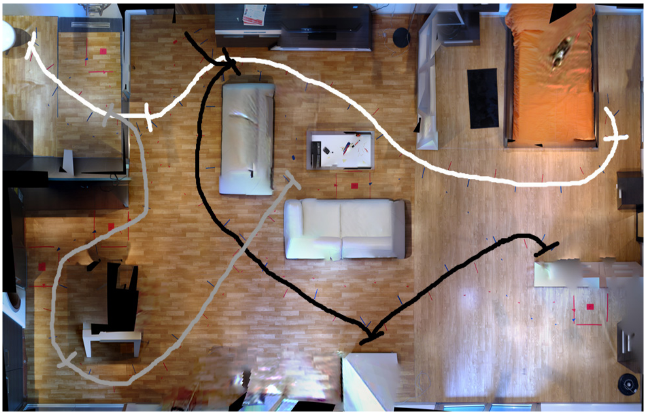

Fig. 2. The three different paths: path 1 (grey line), path 2 (white line), and path 3 (black line)

3 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. 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 score for this phase 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:

$$\text{AS} = \begin{cases} 0 & \text{if } P > 4 \\ 10 & \text{if } P \leq 0.5 \\ \text{4*(0.5-P)+10} & \text{if } 0.5 < P \leq 2 \\ \text{2*(4-P)} & \text{if } 2 < P \leq 4 \end{cases}$$

The score has been computed as the mean of the three scores obtained by each path.

The overall accuracy score has been computed as the mean of the two phases.

Installation Complexity (weight: 20%): a measure of the effort required to install the AAL localization system in a flat, measured by the EC as the total number of man-minutes of work needed to complete the installation. 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:

$$\text{ICS} = \begin{cases} 10 & \text{if } T * N \leq 10 \\ \text{10 * (60-T*N) / 50} & \text{if } 10 < T * N \leq 60 \\ 0 & \text{if } T * N > 60 \end{cases}$$

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

Availability (weight: 15%): fraction of time the localization system was active and responsive. The availability A 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 double of the test duration in seconds. The values of availability AvS has been translated into a score (ranging from 0 to 10) according to the following formula:

$$AvS = 10 * A$$

Integrability into AAL Systems (weight: 10%): The score ranging from 0 to 10 was given by the EC according with the following list:

- 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.

4 Reference Localization System

The reference localization system is essential to measure the accuracy of the competing systems. In fact, the *accuracy* is defined as a statistic associated to the distance from the real position of the user and the estimated one, and the real position must be reliable and consistent. For this reason, the reference localization system was composed by predetermined coordinates of the paths followed by the actor during the competition. As shown in Figure 3, the Living Lab's floor was covered with red and blue marks (for the right and left foot, respectively) that show where the actor had to step on. The synchronization between the steps and the evaluation tool was guaranteed by a digital metronome that indicated the right cadence (one beep one step). In this way we guaranteed that the actor repeated the same paths at the same speed for every competitor.

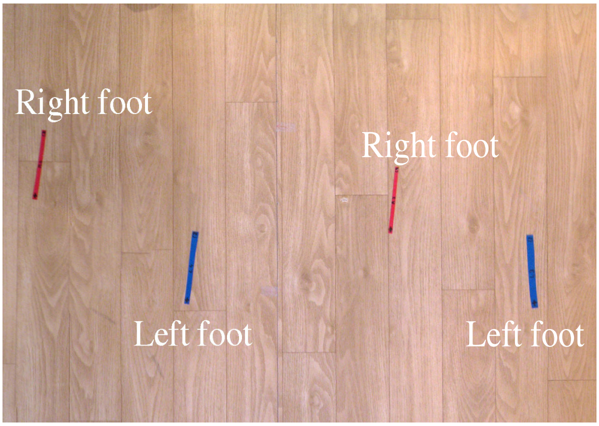

Fig. 3. The reference localization system: the blue marks are related to the left foot while the red ones are related to the right foot

5 Datasets Gathering

During the competition, we collected a large amount of data that are useful as benchmarks to researcher communities who can simulate and test their solutions. The benchmarks collected during EvAAL are particularly interesting as they were collected in a realistic environment with little or no prior preparation.

During the competition universAAL [3] has been used as the software platform to send and store localization events. Competitors were provided with a full running universAAL environment, in addition to a couple of competition specific modules that were developed. These packages include: the competition ontology, as universAAL is a semantic based framework, we had to specify the common semantics for all test runs; and a context provision module, which will make use of the specific semantics to provide location context events. Competitors had two alternatives to interface with the latter module, the primary option being a Java interface that could be imported from any Java project (see figure 4), the second being a simple TCP socket interface with a standard command line like protocol, which could be used by any implementation method used.

The use of universAAL enabled competitors to communicate with the local server, by means of simple localization context events, without any special configuration. UniversAAL will resolve problems like node discovering and inter node communication. In this first competition only basic features of universAAL were used. Based on experience, further competitions will employ broader use of universAAL features.

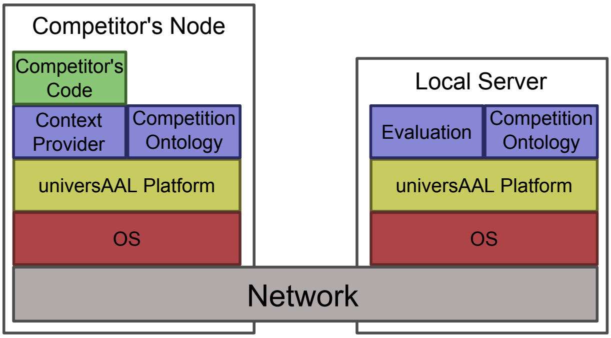

Fig. 4. Communication stack between competitor's node and the local server. Competitors are provided with all modules, even a template and sample of competitor's code.

Locally in the Living Lab a server program has been prepared as a receiver of the contextual information (module labeled Evaluation in figure 4). The local server was in charge of synchronizing the benchmarks and computing the availability and accuracy metrics in real time. The synchronization with the actor's position is also done by the local server namely by the control of the digital metronome. This synchronization is critical for computing of evaluation metrics since these are based on real time information like availability or jitter and correlate with the current position in the reference system. Also in real time the local server displayed a visual comparison of the position provided by the competitors and the real data. This display is used to detect problems in the communication.

Both on the competitors' computer and on the local server logs were produced in a human readable textual format. Although the competition data is also recorded by the local server, the software provided to competitors logged the sent data too, in case of unforeseeable communication problems. The information collected thereby can be resumed as follows:

Competitors' logs:

- Instant of time when the event has been sent to the platform
- Instant of time when the event has been estimated by the competitor
- Area of interest detected by the competitor
- Position (x axis) measured in meters
- Position (y axis) measured in meters

Server logs also include:

- Date and time of the benchmark
- Time difference between competitor's clock and server clock (measured manually at every benchmark)
- Competitor's name
- Benchmark name
- Jitter values
  - min jitter value
  - max jitter value
  - expected period of localization events
- Every localization event is logged with
  - timestamp
  - if it has been accepted or discarded by the jitter algorithm
  - expected coordinates (x,y)
  - expected area of interest

The complete data set has been also rearranged in a spread sheet that computes metrics and shows significant graphs about the actual and estimated position for every benchmark.

6 Methodology

The methodology followed to setting up the competition can be divided into three phases, namely preparation, execution and spreading knowledge.

Preparation. In this initial phase the benchmark and the evaluation criteria have been selected. This phase is probably the most delicate one, in fact, the choice of a benchmark/evaluation criteria with respect to another can lead to a different final result. Both, the benchmarks and the evaluation criteria, have been chosen according with the selected AAL scenario (Section 2 and 3). Moreover, the evaluation software that will be used during the competition is developed and tested. During this phase a call for competition has been spread, receiving 10 submissions where each competitor described its localization system. Through a peer-review process, 6 of the 10 proposed localization systems have been selected for the competition.

Execution. We defined a protocol that was followed for each competitor, and a member of the EC was responsible of making sure that all the steps are properly followed. The protocol consists of 17 steps:

1. When the competitor arrives the EC chair explains the protocol
2. The carpet of the Living Lab was covered before the competitor's arrival. This guarantee that all the competitors don't know the chosen paths before that the installation of their devices is completed.
3. All the members of the EC should be present observing that the rules of the competition are satisfied.
4. In order to produce the metadata for the datasets we measure the position of each device deployed in the Living Lab.
5. We measure the time necessary to deploy the localization system (installation time) in order to evaluate the ICS.
6. We unfold the carpet of the Living Lab
7. The integration between the competitor's software and the evaluation tool is performed
8. Only the actor will be inside the Living Lab during the evaluation phase, and the competitor as well as the EC will be outside.
9. All the evaluation phases, composed by the benchmarks defined in Section 2 will be recorded
10. The evaluation phase start and the evaluation tool will produce the scores relative to the accuracy and availability. These scores will be given follow the evaluation criteria described in Section 3. Moreover, during the evaluation phase, a real time graph has been produced that indicated the path followed by the actor and the position estimated by the localization systems (figure 6), and the estimated area of interest (figure 5).
11. The competitor will be interviewed regarding the integrability aspects (Section 3)
12. The EC will give the score about integrability (Section 3)
13. Each member of the EC give its own score about the user acceptance (Section 3)
14. The final score has been computed exploiting all the computed scores
15. The competitors answer to the questionnaire about the organization of the EvAAL competition
16. In order to collect the datasets the competitor gives to the organizer the intermediate data produced during the evaluation phase.
17. We fold the carpet of the Living Lab hiding, to the next competitors, the paths that the actor will perform during the evaluation phase

Spreading Knowledge. The last phase is related to the organization of a dedicated workshop where the proposed localization solutions are presented. The reason for separating the two phases (execution and spreading knowledge) was that, during the competition, the competitors where admitted one by one to

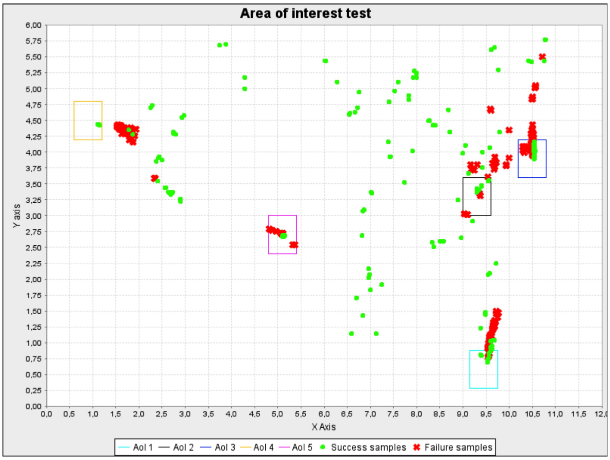

Fig. 5. Real-time graph of the actual and estimated Area of Interest

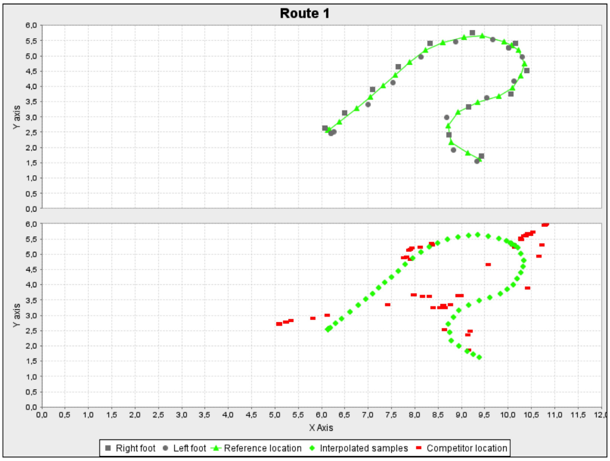

Fig. 6. Real-time graph of the actual (green points) and estimated coordinates (red points)

the living lab: each competitor was given a three hour time slot during which the system was installed and tested, and competitors did not have the opportunity to meet and discuss. The results obtained (datasets, description of the systems used by the competitors, benchmarks, Toolkit development, etc..) where presented later at the EvAAL workshop where competitors were also invited to describe their systems. The AAL Forum was chosen as hosting conference for the EvAAL Workshop because it is a major, annual conference of the Ambient Assisted Living Joint Programme, it has a large audience interested in AAL, and it gives a considerable attention to the most recent EU initiatives. For these reasons, it may provide to EvAAL the appropriate visibility in the scientific and industrial communities working on AAL.

7 Results

At the Ciami Living lab 6 teams challenged themselves at the competition, namely n-core Polaris (from the University of Salamanca) [8], AIT (from Austrian Institute of Technology) [9], iLoc (from Stuttgart University of Applied Sciences and iHomeLab at Lucerne University of Applied Sciences) [10], OwlPS (from University of Franche-Comte) [11], GEDES-UGR (from University of Granada) [12], and SNTUmicro (from Sevastopol National Technical University) [13]. Tables 1 and 2 summarize the scores of the different competitors. In particular, the n-Core system reached a best overall score, 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 a high score in all the metrics, while AIT and iLoc obtained low scores for availability and installation complexity, respectively.

Table 1. The measurable metrics: accuracy, availability and installation complexity

Competitor	Accuracy	Availability	Installation Complexity
n-Core [8]	5,96	9,88	10
AIT [9]	8,45	1,37	6,8
iLoc [10]	7,80	9,39	0
OwlPS [11]	1,37	9,43	8,5
GEDES-UGR [12]	1,81	9,02	0
SNTUmicro [13]	0	0	10

8 Lesson Learned

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.

Table 2. The final scores and the jury decision scores (user acceptance and integrability)

Competitor	User Acceptance	Integrability in AAL	Final score
n-Core [8]	7.6	6.5	7.14
AIT [9]	6,88	8,5	5,90
iLoc [10]	5,88	4,5	4,98
OwlPS [11]	6,5	1	4,85
GEDES-UGR [12]	6	10	4,00
SNTUmicro [13]	4,38	3	3,17

The *Preparation* phase is the most delicate one, in fact the choice of the benchmarks and evaluation criteria for AAL applications is essential for the organization of the competition. These choices could be open not only to the organization committee but also to the competitors, that can give a quick feedback to the organizers on the substance of the proposal.

The *Execution* phase is the most important one, that will constitute the success or not of the competition. We highlighted that, in order to have success, the hosting structure (the living lab in our case) must guarantee the complete execution of the competition, and the integration between the competitor's software and the evaluation tools. To do that the structure must be able to resolve all the possible issues and to spare no effort in the overall organization.

The main trouble of the *Spreading knowledge* phase is the choice of an appropriate hosting conference able to offer logistics and publication facilities, a large audience interested in AAL, and a considerable attention to the most recent AAL initiatives.

9 Conclusions

Feedbacks from competitors and workshop audience were encouraging, for this reason we are currently 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. We conclude with our warm inviting everybody to help us make EvAAL a stable and widely recognized event for AAL. Further reading about the organization aspects of the competition are available on the official EvAAL website [14].

References

1. Connelly, K., Siek, K., Mulder, I., Neely, S., Stevenson, G., Kray, C.: Evaluating pervasive and ubiquitous systems. IEEE Pervasive Computing 7, 85–88 (2008)
2. AALOA association, `http://www.aaloa.org`

3. EU FP7 universAAL project, `http://www.universAAL.org`
4. Trading Agent Competition (2010), `http://www.sics.se/tac/`
5. DARPA Grand Challenge (2007), `http://www.darpa.mil/grandchallenge/index.asp`
6. CIAmI Living Lab., `http://www.ciami.es/valencia/`
7. AAL Forum, `http://www.aalforum.eu`
8. n-Core, `http://n-core.info/`
9. Fuxreiter, T., Mayer, C., Hanke, S., Gira, M., Sili, M., Kropf, J.: A modular platform for event recognition in smart homes. In: 2010 12th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), pp. 1–6 (July 2010)
10. Knauth, S., Jost, C., Klapproth, A.: iLoc: a localisation system for visitor tracking and guidance. In: Proceedings of the Embedded World Conference, Nuremberg, Germany (March 2009)
11. Cypriani, M., Lassabe, F., Canalda, P., Spies, F.: Wi-Fi-based indoor positioning: Basic techniques, hybrid algorithms and open software platform. In: 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–10 (September 2010)
12. Ruiz-Lòpez, T., Garrido, J.L., Rodrìguez-Domìnguez, C., Noguera, M.: Sherlock: A Hybrid, Adaptive Positioning Servicebased on Standard Technologies. In: AAL Forum Proceedings (September 2011) (to appear)
13. Shirokov, I.B., Ponyatenko, A., Kulish, O.: The measurement of angle-of-arrival of microwave in a task of precision landing of aircraft. In: Electromagnetics Research Symposium, July 26, Cambridge, USA, pp. 175–181 (2008)
14. EvAAL web site, `http://evaal.aaloa.org/`