SUMMARY:

Software Testing Report

Complex Project:

PN-III-P1-1.2-PCCDI-2017-0917

Component Project:

P2 - Efficient communication based on smart devices in interactive in-vehicle augmented reality scenarios

Partners:

Ovidius University of Constanța

Ștefan cel Mare University of Suceava

Authors:

Ovidius University of Constanța
Prof.univ.dr. Dorin Mircea POPOVICI
Conf.univ.dr. Dragoş-Florin SBURLAN
Conf.univ.dr. Crenguța Mădălina PUCHIANU
Lect.univ.dr. Elena BĂUTU

Ştefan cel Mare University of Suceava
Prof.dr.ing. Radu-Daniel VATAVU
Prof.dr.ing. Ştefan-Gheorghe PENTIUC
Conf. univ. dr.ing. Ovidiu-Andrei SCHIPOR

Table of Content

1. Introduction

This document presents the summary of the software tests performed with the **Euphoria (Event-based Unified Platform for HeteRogeneous and Aynchronous Interactions)** architecture – a new software architecture design and implementation that enables prototyping and evaluation of flexible, asynchronous interactions between users, personal devices, and public installations, systems, and services within smart environments of all kinds, e.g. the in-vehicle augmented reality system from component project P2. We conducted a controlled scientific experiment with independent variables, dependent variables and scientific hypotheses, following a well specified experimental design with repeated measurements.

2. Experimental Design

The key aspect we were interested in is the temporal performance of event processing in *Euphoria*. We evaluated this performance using the REQUEST-RESPONSE-TIME dependent variable defined as the average time, expressed in milliseconds, required for an event to be processed by the *Euphoria* architecture. Measuring this variable, we are interested in following the entire route of an event, starting from its creation by a PRODUCER until its consumption by a CONSUMER.

The experimental design consists in repeated measurements of the REQUEST-RESPONSE-TIME variable, controlling the following 3 independent variables:

1. MESSAGE-SIZE – interval variable with five experimental possible values, representing message sizes ranging from 64 bytes to 16 Kbytes in a geometric progression with common ratio 4, i.e., 64, 256, 1024, 4096, and 16,384 bytes, respectively. MESSAGE-SIZE represents the amount of information contained by an event created, transmitted, and processed by *Euphoria*. The null hypothesis (H_{i0}) states that the message size does not influence the response time and, hence, there is no relationship among the two.

    The alternative hypothesis (H_{i1}) states that larger messages will be processed in greater amounts of time. Furthermore, we assume that there exists a linear dependence of REQUEST-RESPONSE-TIME as a function of MESSAGE-SIZE.

    H1: Larger event messages will be processed in *Euphoria* with longer request-response times.

    H2: A linear relationship exists between the size of an event message and the time required to process the event.

2. DEVICE-TYPE ordinal variable with three possible values, representing devices with *low-end*, *mid-range*, and *high-end* processing capability; see the "Apparatus" subsection next for details).

3. ENVIRONMENT-COMPLEXITY, interval variable, representing the number of Consumers running requests and receiving responses in *Euphoria* at the same time.

Both MESSAGE-SIZE and ENVIRONMENT-COMPLEXITY are intrinsically linked to the scalability of the *EUPHORIA* architecture. The independent variable DEVICE-TYPE allows testing the flexibility of the architecture, since it allows the experiment to manage a diverse range of device types.

3. Experimental methodology

The size and the content of each message were varied within each experiment, whereas the device types and environment complexities were randomized. Each message generated by a producer was broadcasted by *Euphoria* to all the consumers. This way, the response time could be measured in conditions of maximum network load. We noticed that the results were slightly different in different runs. Hence, we chose to compute and report the average of the values obtained over several trials. Empirically, we observed that the average results remain the same in case of trials that surpass 1,000 repetitions. In the following, we report results obtained over 1,000 repetitions for each possible combination of values of the independent variables MESSAGE-SIZE x DEVICE-TYPE x ENVIRONMENT-COMPLEXITY, resulting in 5 × 3 × 11 × 1,000 = 165,000 experimental trials.

4. Apparatus

The Engine of *Euphoria* ran on a 64-bit Intel Core i7-4510U CPU with 2.60 GHz, 8 GB RAM, and Windows 7. Producers, Emitters, Receivers, Consumers, and simulation modules were implemented on the following mobile devices, representing the three conditions of the DEVICE-TYPE independent variable:

1. Samsung Galaxy S III, a smartphone with 32-bit dual-core Cortex A9 CPU @1GHz and 1GB RAM, running Android 4.1 (our low-end CPU condition). 1GHz CPUs are now common on many wearables, such as smartwatches (e.g., Samsung's Gear Fit 2), while the A9 ARM CPU has been super-seeded by more advanced architectures; see next.

2. Samsung Galaxy Tab 4, a tablet with 64-bit quad-core Cortex A53 CPU @1.4GHz, 1.5GB RAM, running Android 4.4 (our midrange condition). Compared to the A9 architecture, A53 has more CPU cores, higher CPU frequency, support for 64-bit applications, hardware-assisted virtualization, dynamic frequency scaling, and an enhanced Floating-Point Unit. More RAM memory and an improved operating system imply better performance.
3. Acer Aspire V15 Nitro, a notebook with 64-bit quad-core i7-4710 CPU @2.50GHz, 12 GB RAM, running Windows 10. Compared to A9 and A53, the i7 CPU architecture relies on an advanced trigate transistor technology that allows low power consumption and high processing speeds. It also offers hyper-threading technology and dedicated instruction sets, such as FMA (Fused Multiply-Add) and AVX (Advanced Vector Extensions).

5. Experimental Results

First, we studied the effect of the variable MESSAGE-SIZE on the variable REQUEST-RESPONSE-TIME for each value of the independent variable DEVICE-TYPE. On the average, the response time varied between 11ms (for messages of 64B, received by consumers that run on *high-end* devices) to 141ms (for messages of 16KB received by consumers that run on *low-end* devices. These results entitle us to reject the null hypothesis H_{10} and to accept the alternative hypothesis H_{1A}: larger event messages are processed with longer request-response times.

The effect of the variable ENVIRONMENT-COMPLEXITY was investigated next, for each DEVICE-TYPE. On the average, the response time varied from 9ms (as was the case for 1 consumer, running on a *high-end* device) to 113ms (the case of messages being broadcasted, simultaneously, to 101 consumers running on *low-end* devices). These results entitle us to reject the null hypothesis H_{40} and to accept the alternative hypothesis H_{4A}. This result also confirms H_{3A} and leads us to reject $H_{30,}$ more powerful and resourceful devices process messages faster.

Our empirical results show that *Euphoria* can process messages with a temporal performance that can be perceived as instantaneous response by users: under 150ms average request-response time for the slowest device (A9 CPU) that processed the largest messages (64 KB) in the highest density condition (101 consumers) evaluated in our simulations.

We mention that the *Euphoria* architecture was not tested in critical situations, that involve real time responses, as it was not our main goal for the moment.

6. Application Scenarios for in-vehicle usage. Preliminary evaluation of the perceived usability.

During this phase of the project, we explored various aspects of data acquisition, feedback types (visual, audio, haptic) for various in-vehicle interaction and usage scenarios. To this end, we were interested in the possibility of augmenting the command and control system of a vehicle using gestures executed by the driver. To achieve this objective, we proposed an architecture composed of

- a software module for user gesture detection, for which we tested both the Leap Motion device and the Myo bracelet

- a visualization module for the most probable selection, using a mini-video projector

- a software module that offers the connection between the gesture of the user and the execution of a command, for which we implemented a list of options to be selected by a finger movement.

We tested the usability of these developments regarding the utility, the comprehensibility, ease of learning and operability. We used an evaluation methodology based on measuring the user perception on a Likert scale with 5 items (1 – Total disagreement to 5 – Total agreement). The charts presented in Figure 1 illustrate the answers of the survey participants to two questions selected from the survey.

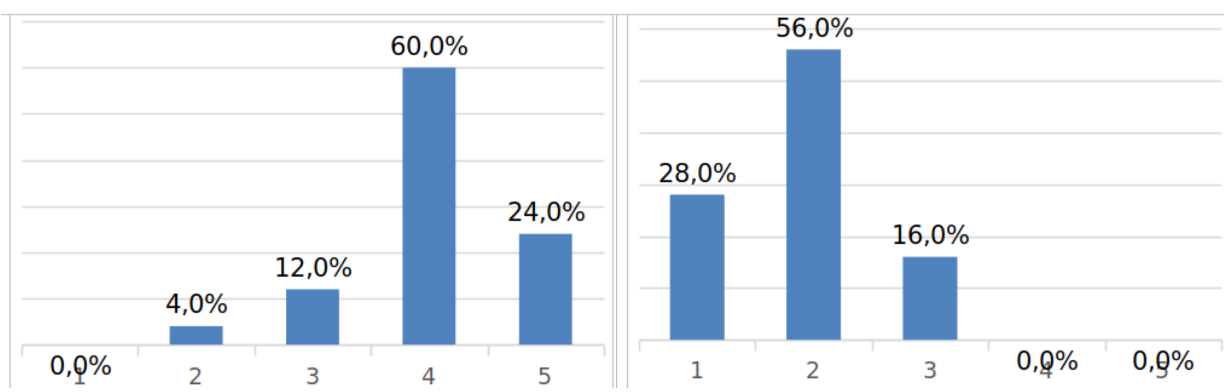

Figure 1. Results from the usability survey for the Leap Motion & Myo prototype application. Left: the perceived comfort, obtained as a direct response from the users for the item "I felt comfortable using the app.". Right: the ease of perceiving the projected information, obtained as responses to the item "The information projected by the visualization module is hard to see."

The *GenericProducer* and *GenericConsumer* apps are, at this point, in the prototype stage. Nevertheless, we were interested in the measuring the satisfaction degree of users. Most of the users that participated to the usability survey have extended experience with mobile apps, including Augmented Reality apps. The 10th item in the survey is “I enjoyed using the app.” The users’ answers agree with this sentence (see Figure 2). Also, users considered the app to be interesting.

The readability of information was positively appreciated by the users; 96% of the users responded with 4 or 5 to question nº 9 “I consider the displayed symbols and words were easily readable.” The speed with which the application allows access to received information is also appreciated by the users that participated to the survey: 93% of them responded 4 or 5 to question 13, “Using the app, I have fast access to the received information.”

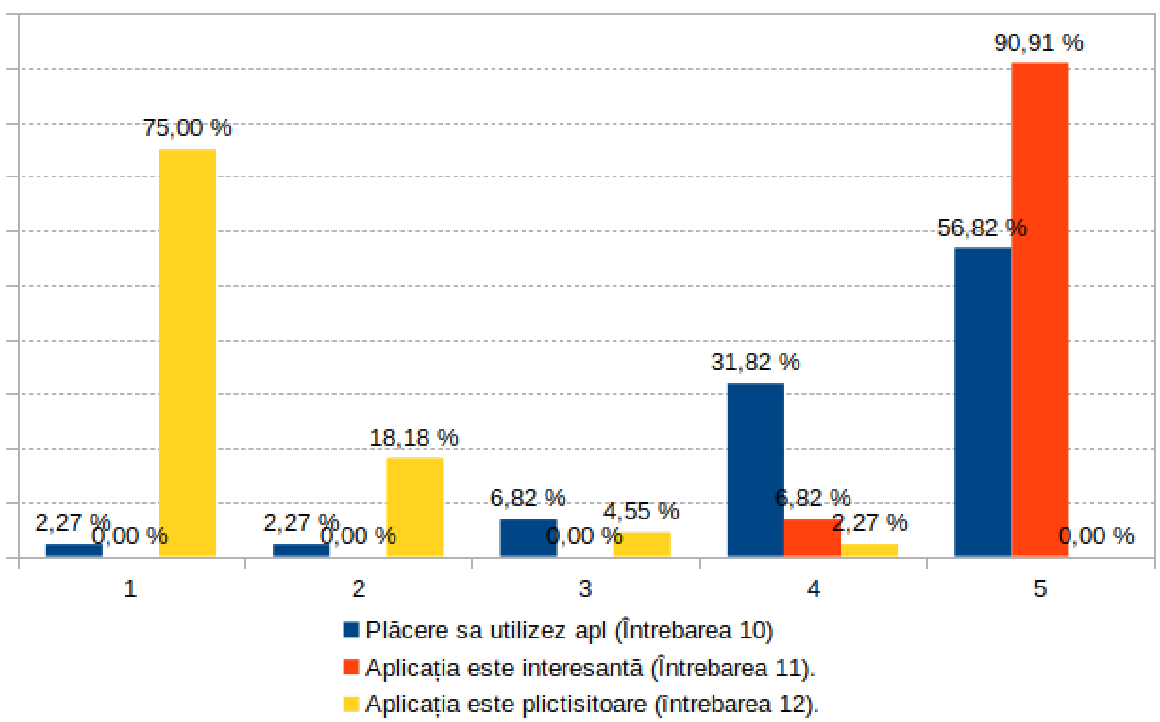

Figure 2. Satisfaction degree of the surveyed users (Question 10: Using the app was pleasant. Question 11: The app is interesting. Question 12: The app is boring.)

We identified several future implementation directions that use the producer/consumer modules developed during this phase of the project. For example, with respect to the display of information received from the producers, our main interest was to attain high visibility, regardless of the environmental conditions. Also, we were interested to avoid the distraction of the users' attention, both for the users generally involved in the traffic, and, particularly, for the automobile driver.

Various technical solutions were considered, beginning with AR helmets (Figure 3a), mini-projectors (Figures 3b, 3c), smartphones/tablets or HUD devices (Figures 3d, 3e, 3f), during daytime (Figures 3d, 3f) as well as during nighttime (Figures 3g, 3h). These solutions and technologies shall be considered in the next phases of the project.

We consider that the results of the evaluation survey are encouraging, considering the high level of usability and acceptability appreciated by the users surveyed. Nevertheless, we must consider the influence of the characteristics of the surveyed users: they were mostly men, with a high level of experience with mobile apps. Further work will be devoted to quantifying this influence, if any.

7. Conclusions

We tested the *Euphoria* architecture in 165 distinct configurations for each combination of the independent variables MESSAGE-SIZE x DEVICE-TYPE x ENVIRONMENT-COMPLEXITY. To avoid validity threats posed by cache or network usage, we blocked all automatic updates on all 3 devices, such that all experimental trials are based on the same software configuration. Moreover, we randomized these combinations for 1,000 repetitions. The specific characteristics of our experiments enabled us to measure precisely the timely succession of events. Nevertheless, although we intended to be as objective as possible, our experiments may suffer from the choice of values for the DEVICE-TYPE variable. This can happen because there exists no concept of a generic device that completely characterizes a market share. During the entire experiment, we maintained the same software and hardware configuration for the *Euphoria* architecture, to avoid instrumental errors. Also, the large number of repetitions should compensate for the (unpredictable) influences.

To further strengthen the external validity of our experiments, we provided all information necessary to replicate our results. This includes the software that was developed as a basis of our architecture.

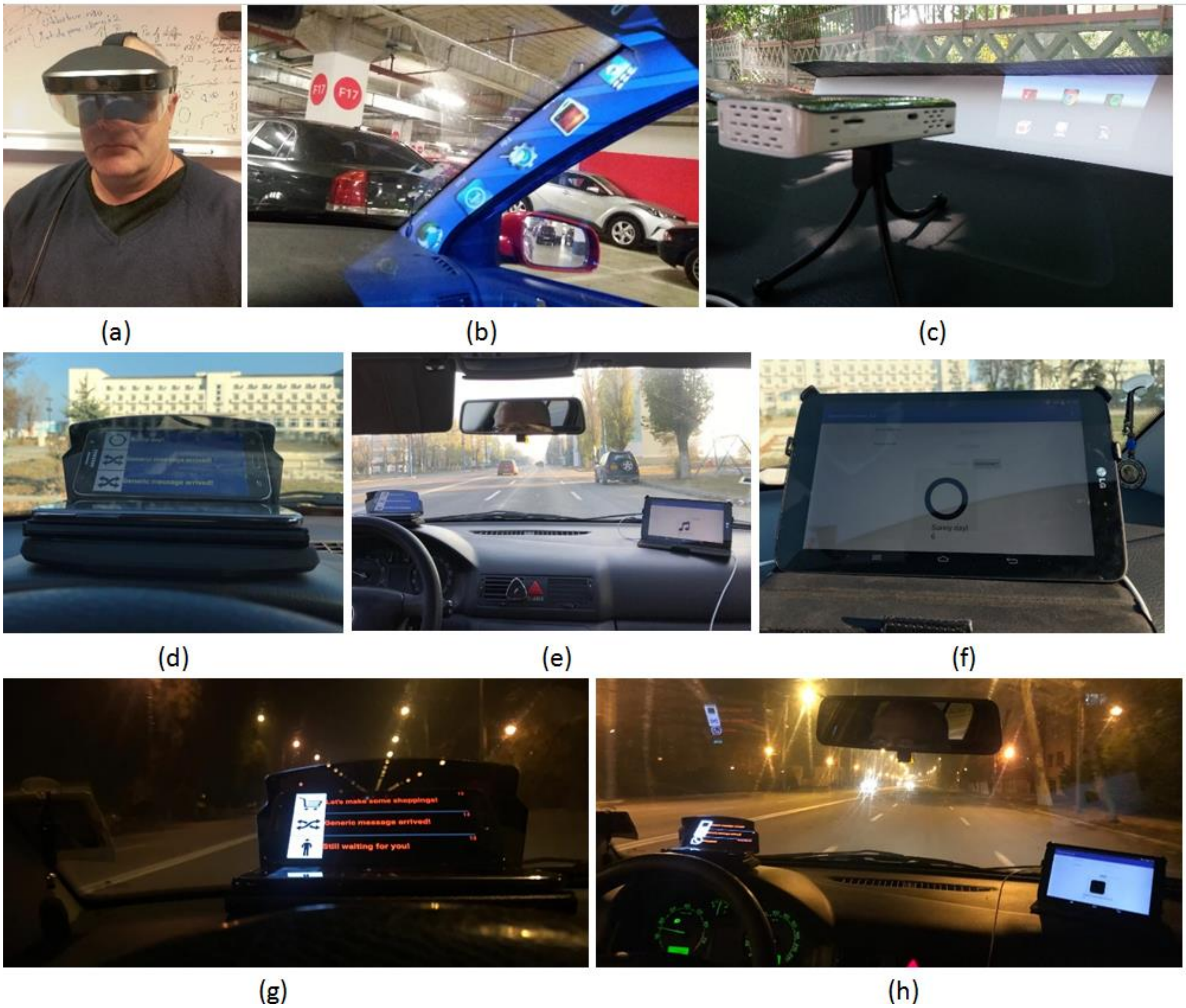

Figure 3. Exploring various configurations for visualizing information inside the vehicle.