Evaluation Report of MobileSage Services 2.0

Authors:
Till Halbach, Trenton Schulz

Version: 1.0
Date: 2014-06-06
Dissemination Level: Public

Abstract

This deliverable reports the third and final user evaluation of the MobileSage prototypes Help-on-Demand and Content Management Systems. It also sums up all evaluations carried out in the project period.

A small-scale evaluation was carried out in Norway in Spring 2014. Although participants had issues interacting with the Content Management System, they were able to create video content and enter the information into the system. When testing the Help-on-Demand mobile application, participants had questions about the terminology used in the interface. Participants wanted to have both systems available for further testing and to put into use. They especially pointed out that they wanted to have easier access to the mobile application for installing on other phones.

Table of Contents

Introduction

This section details the background and scope of this Deliverable.

Background

MobileSage is an Ambient Assisted Living project (AAL) [1]. The main objective in MobileSage is the development of two service applications [2]. The first is a mobile client application, called *Help on Demand Service [3]*, and the second is a Web server application, here referred to as *Content Management Service*.

This document is a deliverable of Work Package 4 of the MobileSage Project, labeled Testing and Evaluation. Other deliverables from this Work Package are as follows:

•D4.1 Evaluation report of Content Management Service,

•D4.2 Evaluation report of Help-on-Demand Service version 0.9 and changes, and

•D4.3 Evaluation report of Help-on-Demand Service version 1.0 and changes.

Scope of the Deliverable

This document reports on the findings from the project's Workpackage 4, labeled Testing and Evaluation, and also summarizes the overall evaluation findings from the project.

1Introduction

The MobileSage prototype was developed in three major iterations, where the release of a software deliverable, dubbed Beta, V1.0, and V2.0, marked the end of an iteration. Each release was evaluated involving end users. The first and the second evaluations were carried out in Norway, Romania, and Spain, while the last evaluation was conducted in Norway only. In total, around 70 informants were part of the evaluations in the three countries.

The subsequent sections summarize the findings from each iteration.

2First iteration: Beta evaluation

The Beta evaluation consisted of different activities for each country. Romania used a workshop where participants were introduced to MobileSage. Spain used a focus group to present the MobileSage

concept with an acceptance test. Norway performed a user evaluation of the HoD app. Full information is included in D4.2.

2.1Beta setup

2.1.1Romania

In Romania, the workshop consisted of a group of 10 elderly participants. Five were female with compensated audio-visual impairment. The other was a group of elderly with memory disturbances consisting of four males and one female. None of the participants used smartphones. Four received text messages and two also send text messages.

The session consisted of two parts. First, a presentation of the project and the first HoD prototype. Discussion was recorded for analysis. Second, each participant was given time to use a Samsung Galaxy Note 2 phone with the HoD app installed. Participants were encouraged to explore the functions of the phone with support from one of the leaders of the focus group. They were also encouraged to use the HoD app, scan a code for the coffee machine to get a video or graphic version. Finally, participants filled in a questionnaire.

2.1.2Spain

The evaluation in Spain consisted of a focus group held at Telefonica in Barcelona with ten retired participants between the ages of 65 and 70 years. Some participants had smartphones, but only used the normal mobile phone functionality of calling and text messages. Some did take pictures and listen to music on their phones.

The focus group was based on an acceptance test based on the System Usability Scale (SUS). Participants were first asked about their feelings on smartphones and mobile applications to establish a baseline. This was followed by a demonstration of smartphones, the MobileSage service, and how it could help them in their day-to-day routines. Participants were also given the chance to work with the HoD app.

The participants were then also presented with a two scenarios. One for traveling in a city where the traveler was not a native speaker, and the other of trying to create a cup of coffee. They were also shown possible videos that would be delivered by the HoD app. After discussion, participants filled in the acceptance test again.

2.1.3Norway

The Beta evaluation in Norway consisted of a travel situation at a subway station in Oslo, where participants used the Beta prototype to find help with getting to a subway station, finding a ticket machine, buying and validating a ticket, and choosing the correct platform. We created content for all of these scenarios, with a minimum of audio and video for each. All but one of the scenarios

had captioned video, and some had a textual modality in addition. Each of audio, text, and video was done both in Norwegian and in English to allow users to choose an additional content language. Several NFC tags were then written for each of the scenarios and used as a way of obtaining the desired information. Testing of QR codes was postponed to a later evaluation due to its unreliability in the Beta version of the app. We tested on two smartphones with an Android OS 4.1 and 2 different screen sizes without any discernible difference in the results.

Eight participants were recruited for the evaluation. They were between 65 and 76 years old, and four of them had no experience with smartphones; however, they were somewhat experienced with computers. A few of them knew the location to a certain extent. A session consisted of a short introduction to the MobileSage idea, followed by a brief interview concerning their experience with mobile phones. The application was then demonstrated, after which the informants would work on the tasks. One evaluator took notes, while the other would guide the participant to make sure a task wasn't forgotten. After completing all of them, there was a short follow-up interview about the service and the participant's experience about it.

2.2Beta results

2.2.1Romania

Many of the participants had difficulties performing tasks because of the unfamiliarity with the smartphone. There were many comments about use of smartphones in general. Participants had trouble finding the application icon and wanted it bigger. They also felt that the phone should be simplified or personalized for them and having human help in instructing them in its use.

Looking at the HoD app specifically, the biggest issue was using the scanning function. No one considered the HoD's scan function difficult, but felt they needed help to learn how to use it. Many were unsure if they would use the function outside of the coffee machine scenario.

After scanning the QR code, there was an issue in understanding the information that was presented. Most preferred getting the information in the form of video instead of only graphics. But, everyone felt that the content needed to be improved. They also wanted to be notified when there were issues with the connection.

Overall, they found the MobileSage idea interesting and felt that it was worthwhile learning.

2.2.2Spain

The acceptance test is based on a scale from 1 to 5 with higher scores being better. For the initial round, the acceptance score was

1.87. After participants had a demonstration of smartphones, experienced the scenarios, and tried out the phones, the score became 2.20. This is an increase from when they started, but is still not a very high score. It did show that explaining the use cases for smartphones and showing how it can improve everyday experiences can interest the elderly, but the cases may need to be more compelling.

2.2.3Norway

In the first task, the participants had to create a profile that matched their preferences for text size, language, and types of media they wanted to receive help in. The users understood the concept of several content languages, and the majority (75%) added English to their profile. The user-specific media types ranged from a single one to the entire range, where captioned video, i.e., the richest media type, was chosen most often. The majority (90%) of test persons checked 4-5 media types including audio, even though some participants said they would avoid wearing earbuds or headphones. Text was never requested. Choosing video and captioned video was inconsistent, hinting on a potential misunderstanding of the user concerning the meaning of ”captioning”. It was recommended to improve the description of media types or show brief examples of them. All participants but one expressed that making the profile was sufficiently easy.

The second task concerned navigation, where the participants had to get from their current location to the nearest subway station. All were able to enter the information needed, but the phone's positioning worked unreliable, sometimes placing the participant a block further south or facing the wrong direction. This issue sorted itself out when walking to the location.

The next task dealt with getting help at the ticket machine. Two participants were not able to finish this task due to a technical issue that caused no results to be returned from the CMS, which was corrected subsequently. All others succeeded with using NFC tags or by manually searching for information about where the ticket machine was, how to purchase a ticket, how to validate the ticket, and which platform they had to go to. Though only one was familiar with the technology, two had heard about it, and the rest were unaware of what it was, all really liked the technology and experienced it as easy to use.

One problem encountered was the effect the environment had on the signal strength in the phones. While above ground, it was possible to get video and audio without any issues, and the selected item would show up almost instantly. Yet, underground in the subway station, it became very troublesome for the phone to contact the content server. The main reason for this is that the only connections that are currently available in this particular station are so-called Edge (2G) connections, which are much slower compared

to a 3G connection, and also very latent. This was no big issue when retrieving, say, the results list. Participants had to wait a long time, though, if they wanted to watch a video. The audio fared a little better, but downloading would not always complete. Sometimes, the application on the phone would simply give up and it would be necessary to download the audio or video from the beginning. Most participants noted that it took a while to get the information in this case. With the continuing widespread of 2G connections in many countries, it is recommended to produce at least one version of low-resolution low-bitrate content, and to use techniques that increase the responsiveness of media players, such as media streaming.

No users complained about the size, resolution, quality, frame rate, or length of the video. Some participants noted that the font used for the captions indeed was sufficiently large and easy to read. There was only one instance where people commented on unclear information, where a video showed an unreadable display on a ticket machine.

All participants felt that a help-on-demand system was something that would be useful for them. One even claimed that she was scared of using the ticket machine and always went to a counter instead, but now she would continue to use the machine since she felt confident to manage buying a ticket based on the app and the provided instructions. Concerning potential improvements, the most popular suggestions were a shorter response time for videos (when in the subway station) and dynamic information, such as time schedules. Those familiar with mobile applications suggested to include MobileSage's functionality in the public transport provider's current smartphone application.

3Second iteration: V1.0 evaluation

Each country had its own set up for the evaluations. Romania had a workshop with potential users of the HoD app. Spain conducted interviews with potential users. Norway went with a standard user evaluation of the HoD app. The full report is included in D4.3.

3.1V1.0 setup

3.1.1Romania

The V1.0 evaluation in Romania was held at AAIF's Clinic of Memory Diseases in Bucharest and used two indoor mobility scenarios that better matched this group of users. The goals were to see the improvements in user acceptance of the system since the Beta evaluation, see how the indoor mobility scenarios were accepted, and gather suggestions for other improvements. There were ten participants for this evaluation, aged 56 to 80 with the media age of 61. Five participants had compensated audio-visual impairments,

and five had mild memory disturbances. All participants used non-smartphones. Four participants received text messages, and two of them sent text messages from their phones.

The indoor mobility scenarios concerned Marta, who is 70-years-old, lives alone, and had similar impairments as the participants. She is able to accomplish most tasks on her own, but appreciates any help. She also has a son who is traveling abroad and she likes to keep in touch with him.

The first scenario involves cooking a recipe from a foreign language cookbook. Marta's son has put a QR code in the book that can be scanned by the HoD app to get a translation provided by her son. The second scenario involved being able to look at pictures from her son's vacation and listen to music with help of the HoD app.

Participants were given a presentation of the project and the HoD. They then had 10–15 minutes to try out the smartphone, a Samsung Galaxy Note 2. They then explored the app itself and performed the tasks related to the scenarios. Afterwards, they filled out a questionnaire.

3.1.2Spain

The evaluation in Spain focused on a tourism scenario and the use of QR codes. The participants were ten subjects that were all over 65 and interviewed in their homes.

The scenario focused on an elderly couple from Madrid that was visiting Barcelona for the weekend. They used the HoD app and special HELP QR codes to find out different types of information including directions to the museum, information at the museum, setting up Wi-Fi, how to use a ticketing machine, daily menus at a restaurant, and booking a taxi. All the information is provided in Spanish.

The testing procedure consisted of the interviewer explaining the problems that MobileSage is trying to solve and introducing the scenario. Participants could then experiment with the smartphone and work through different parts of the scenario. Questions were asked by the interviewer along the way. After the scenario was complete, the participant filled in modified SUS questionnaire.

3.1.3Norway

In the V1.0 evaluation all of the attention was given to the user experience when using the HoD app. The Norwegian evaluation involved 10 informants from the local senior user group, aged 67 to 83, from both genders, all with varying ICT and mobile-phone experience (though none were novice ICT users), and a few with a mother tongue different from Norwegian. All informants received a small financial gratuity for the participation. Two Android phones (Galaxy Nexus and Nexus S) were used in the tests.

The scenario's focus was on matters not tested in the last trials and emphasized multilingual content, the concept of steps, and QR codes. The following tourist situation was considered: A foreign visitor to Norway arrives at the tourist information in Oslo to catch an eye on a poster mentioning the famous Kon-Tiki Museum. The poster provides both an NFC tag and a QR code, either of which is scanned by the user through the MobileSage app, upon which several pieces of information are presented: About the museum and how to get there from the user's current position, how to buy a ticket at the nearby machine, how to find the proper bus stop, when the next bus is arriving, and when to get off the bus while riding to the museum.

Most of these five steps were presented in multiple modalities, such video, audio, and HTML, others were available just in a single media type. The latter two steps showed dynamic content (real-time data) from the servers of the municipality's transport company Ruter, namely the expected duration of the waiting time for the next bus, and the expected duration of the arrival of the bus at the proper bus stop. This was achieved by HTML redirects from the content provided by the CMS to Ruter's server.

The informants were first briefly introduced to the MobileSage idea in general, and scanning of NFC/QR in particular. After that they went to the nearby tourist information where they found a poster as described above. During the session the participants were simply watched as they went through the steps of traveling to the museum. In case of problems we would assist the user with clarification and also give technical aid. Throughout a single trial, the participant moved from the poster to the ticket machine, and further on to the bus stop, while the last step (the bus ride to the museum) was simulated only for practical reasons. After that, the user was interrogated about their user experience and had to fill out a brief questionnaire to gather their opinion.

3.2V1.0 results

3.2.1Romania

Many of the issues that participants in Romania had were related to the their lack of familiarity with the smartphone. Most participants felt that the start screen for launching apps was difficult and expressed that they needed to become more familiar with the phone. Nearly all participants were reluctant in changing settings on the phone. They expressed a need to have some training with a human assistant to understand the settings. This was less of an issue when choosing language. A couple of participants had trouble scanning the NFC tags. Participants felt that the app icon on the start page should be bigger.

Outside of these issues, participants liked the clarity of the content and most preferred audio and video followed by text. Many felt that

they would use most of the features of the HoD frequently or when needed, but a couple of them would use it more frequently after they had learned how to use it. Participants liked the scan functionality, but preferred QR codes to NFC as NFC required them to find the proper position for the phone to scan the tag.

Overall, all participants felt that it was worthwhile to learn the MobileSage service, and that it had high utility. They felt that most used functions should be at the top of the screen and that unneeded or unused features should be hidden from them. But, participants were concerned that they needed to learn how to use the phone and app. Having some sort of human assistance to learn the app should be considered.

3.2.2Spain

The SUS (System Usability Score) was used in both the Beta evaluation and the 1.0 evaluation in Spain. The SUS delivers a value between 1 to 5 to measure acceptance with higher scores being better. The combined score for all participants in the evaluation was 4.23. This is a 92.3% increase from the Beta evaluation. This showed that the improvements that were suggested from the previous evaluation had a good effect. All participants found the application simple to use,, consistent, and well-integrated. Those who were not familiar with smartphones needed help in setting up the app and starting to use it, but felt that they could use it afterwards.

Overall, the participants felt that the HoD was a valuable service. Some minor issues included scanning QR codes, but participants found the scan function valuable. The app crashed a couple of times, and this concerned participants. The final version needs to be more robust.

3.2.3Norway

The findings from the Norwegian trials are presented as an excerpt from the CENTRIC conference paper.

Overall, the English MobileSage version was acceptable for the English-speaking testers, even though they commented on several non-translated page elements in the dynamic webpages from the travel company. Integration of services, including the proper communication of the user's language, is key here, besides the mandatory translation of all language strings.

The participants found the prototype in general accessible, but there were several issues related to real-life situations: traffic, crowd, and town noise was a problem when trying to hear the sound of the videos, both in- and outdoor. All participants would use the relatively weak speakers in the smartphones. As one of the participant remarked, “elderly never use headphones, you know.” Here, video captions were useful to the participants. Next, bright outside

sunlight reduced the screen contrast, making it difficult to read what was displayed. Here, automatic adjustment of the display's brightness/contrast and improved displays would help, but this is beyond the scope of this project. Some participants found the text size and also the virtual-keyboard letters too small. However, even though it had been pointed out to participants that they could adjust the settings according to their own preferences, none actually changed the default text size. The size of the keyboard letters could not be changed, and this might be the reason why the users found NFC and QR codes so attractive when searching for information.

The informants all agreed that the ability to customize and personalize media modalities and output was valuable for them and other elderly as well. Due to time constraints, though, this topic was not tested systematically. The fact that no user entered the settings shows on the other hand the necessity for suitable default values, such as captioned videos as default media type as it combines a visual with audio and text.

Regarding adaptivity, the trial observers could for instance notice that the most used functionality was automatically placed in a prominent position in the user interface (on top). However, the participants did not seem to notice. We did collect usage data for each participant, but due to the small duration of each trial, a sufficient amount of data for each participant was never generated. Future work should test out adaptivity in a realistic manner.

The participants had quick access to the content and were all satisfied with the response time. It surely helped to hold the evaluation in an area with a good GSM signal, but contributing to this was also the strategy to switch from plain downloading in the first trial to HTTP pseudo streaming, which shortens the time after which the media player starts playing drastically.

Most informants had heard about QR codes or at least said they had noticed it, but very few knew about NFC, let alone its logo. All participants preferred scanning over text-based search during the trial. Here, nine out of 10 found that NFC was easier to use than QR due to a shorter response time. With QR, many found it cumbersome to find the correct distance and angle between the smartphone camera and the QR code on the poster. In contrast to the beta trial, NFC tag scanning went smoother, mainly because we now carefully had placed the tags with some distance to any metal surfaces.

As opposed to the beta evaluation, the app now rendered the content of the result directly if only one had been found. Most participants preferred this, but were in turn confused when the result consisted of several steps, as showing a step overview had been omitted. Other than that, steps as a concept was well understood and accepted.

As one of the outcomes from the user interviews, tab:sus shows the general user acceptance, measured by means of the System Usability Scale (SUS). The table clearly documents the improvements in user experience on almost all issues as compared to 2012. The largest positive change occurs related to the app's ease of use, with an increase from 3.1 (uncertain) to 4.4 (clear agree) average score. Overall, our participants had a positive view on the MobileSage system and found it useful and relevant. The scale also shows potential for improvement, however, when it comes to opinions concerning the app's ease of use.

4Third iteration: V2.0 evaluation

The third evaluation was only conducted in Norway.

4.1V2.0 setup

In the V2.0 evaluation, we looked at the final version of the CMS to see how well it worked, and the final version of the HoD app to get feedback on changes to the app. It was a limited evaluation held in cooperation with three participants from the local senior user group that had been involved in the project. Due to time constraints, the evaluation was held as a workshop with all three participants using the system and giving comments simultaneously.

The participants were given a short introduction about the CMS and were then asked to create instructional content that they subsequently uploaded to the CMS. Participants chose to create a video in three parts (steps) for heating water in a microwave. They shot the video on one of the phones, copied the videos over to laptop, and proceeded to put them on the CMS.

For the HoD part, many of the tasks were borrowed from the previous evaluation. We gave the informants a short introduction to the HoD, specifically focusing on the media types setting. The participants were then instructed to find a video about tickets by using the scanning functionality after having changed the media type settings to accept only formatted text and plain text.

The session was concluded with a short discussion about the app and the CMS.

4.2V2.0 results

Subsequently, the findings from the Norwegian trials are presented as an excerpt from CENTRIC conference paper.

Concerning the CMS, participants were confused in the beginning about how content was organized and had problems understanding that a record can consist of multiple steps. Related to this is that a user needs to fill in two titles, one for a step, and another for the

entire record. This information was not included on the web page itself and needed to be explained by the trial observers. Once the concept was understood, participants were able to create and add content containing multiple steps.

Participants encountered problems related to the media type of the content to upload. Depending on the file type (MIME type), particular buttons (“Continue”, “Publish”) were shown or hidden. Currently, only a limited number of formats are allowed to be uploaded in the CMS. However, the negative result from the media type check was not communicated to the user who then had to assume that the upload form did not work. Hence, the user needs to be properly informed about each requirement to an input field, and each conducted check. Aside from this, the participants were able to complete all the steps and fill in the content summary, but they could not actually add the content to the CMS.

All participants commented on the necessity of extra information for creating and adding content in terms of tips on how to create videos so the videos would be more instructional. One participant feared they would include too much information into the content. A take-away here may be to provide tools that can provide assistance for proper content generation.

The CMS makes content available for all, which turned out to be of concern for some participants. They commented that users could hesitate to upload information if it was public for everyone, even though the trial observers pointed out that the location information helps to limit exposure for information, and that the majority of all information is designed to be for public access anyhow. It is noted here that it is technically possible to restrict access to a record say with a passphrase, but this of course adds to the complexity of the system.

In the settings of the app, when choosing media types, users were presented with the jargon terms for the types in the database and not a suitable translation in their own language, leading to questions about the meaning of each term. The built-in help in the app does indeed explain these types without using jargon, but this help is not available when choosing types. The conclusion here is that technical jargon should be avoided, and that concise and explaining help should be available where challenges might arise.

Using the different help-on-demand functions in the HoD app, including scanning of QR codes and NFC tags, worked as expected. One phone, however, was a bit large, and users had to move the phone forth and back to get to read the tag. We believe this problem would vanish as the users become used to their phone.

Generally, the participants were excited about the possibilities of the MobileSage service and wanted to use both systems (app and CMS) more.

We are going to address a number of the issues found in the various evaluations before a final version of each system is released. The CMS is currently available online, and the HoD app is offered through Google Play.

References

1 AAL Association in 3rd call for proposals (ed. AAL Association (AAL), 2010).
2 Sánchez, V. in Mobilesage – situated adaptive guidance for the mobile elderly (ed. Sánchez (MobileSage Consortium), 2011).
3. Røssvoll, T.H. User Requirements Specification. MobileSage Deliverable D2.2, (2012).
4. Curescu, L., Anghelache, I. & Røssvoll, T.H. System Requirements Specification for Help-on-Demand Service. MobileSage Deliverable D2.3 v1.1, (2012).