



## SEVENTH FRAMEWORK PROGRAMME

FP7-ICT-2011-1.5 Networked Media and Search Systems  
b) End-to-end Immersive and Interactive Media Technologies

Specific Targeted Research Project

# VENTURI

(FP7-288238)

immersiVe ENhancemEnt of User-woRld Interactions

D2.1.2: Use-cases, application definition and system requirements for STE platform  
for the second year demonstrator

Due date of deliverable: 30-04-2013  
Actual submission date: [07-05-2013]

Start date of project: 01-10-2011  
Duration: 36 months

## Summary of the document

Document Code:	D2.1.2: Use-cases, application definition and system requirements for STE platform for the second year demonstrator
Last modification:	07/05/2013
State:	Ready for Submission
Participant Partner(s):	ST-Italy, STE
Editor & Authors (alphabetically):	Editor: Giulio Urlini, Paul Chippendale Authors: Olivier Pothier, Giulio Urlini
Fragment:	No
Audience:	<input checked="" type="checkbox"/> public <input type="checkbox"/> restricted <input type="checkbox"/> internal
Abstract:	This document contains the description of the use-case for the demonstrator of the second year of the project highlighting the specific requirements for a next generation platform to increase or fulfill the use-case requirements
Keywords:	Use-case, demonstrator, hardware and software requirements
References:	Refer to the corresponding section at the end of the deliverable

## Document Control Page

Version number	V0.3
Date	07-05-2013
Modified by	Paul Chippendale
Comments	Final quality check
Status	<input type="checkbox"/> draft <input checked="" type="checkbox"/> WP leader accepted <input checked="" type="checkbox"/> Technical coordinator accepted <input checked="" type="checkbox"/> Project coordinator accepted
Action requested	<input type="checkbox"/> to be revised by partners involved in the preparation of the deliverable <input type="checkbox"/> for approval of the WP leader <input type="checkbox"/> for approval of the technical coordinator <input type="checkbox"/> for approval of the project coordinator  Deadline for action: 07/05/2013

## Change history

<b>Version number</b>	<b>Date</b>	<b>Changed by</b>	<b>Changes made</b>
0.1	29/04/2013	Giulio Urlini	Integration of the separate documents already reviewed by the partners and the project coordinator of the second year use-case description and the requirements for the next generation platform (D2.2.2)
0.2	03/05/2013	Olivier Pothier, Giulio Urlini	Reviewed by the WP leader. The applied changes are: <ul style="list-style-type: none"> <li>• Modification of deliverable title to reflect the change of scope</li> <li>• Addition of an introduction to explain the change</li> <li>• Addition to requirements list taken from D2.2.2</li> </ul>
0.3	07/05/2013	Paul Chippendale	Final Quality check



Table of Contents

Summary of the document..... 2

Document Control Page ..... 3

Change history..... 3

Executive Summary ..... 7

    Scope ..... 7

    Audience ..... 7

    Summary..... 7

    Structure ..... 7

1. Introduction ..... 7

2. Use-case description: Personal Assistant for the Visually Impaired ..... 8

    2.1. Introduction..... 8

    2.2. Embodiment..... 8

    2.3. Storyboard..... 9

        2.3.1. Phase-1: route from bus stop to shopping mall entrance ..... 9

        2.3.2. Phase-2: route inside the base floor of the shopping mall..... 9

        2.3.3. Phase-3: Lift usage ..... 10

        2.3.4. Phase-4: second navigation in the mall ..... 10

        2.3.5. Phase-5: navigation inside the shop ..... 10

        2.3.6. Phase-6: navigation on the target shelf..... 10

    2.4. Example Maps ..... 11

3. Requirements for the actual device and the future devices ..... 14

    3.1. Hardware functional requirements..... 14

        3.1.1. Camera..... 14

        3.1.2. Connectivity ..... 15

        3.1.3. Sensors..... 15

        3.1.4. Input method ..... 15

        3.1.5. Display..... 15

        3.1.6. Graphics Hardware ..... 15

- 3.1.7. Audio..... 16
- 3.1.8. Power..... 16
- 3.1.9. Autonomous mode..... 16
- 3.1.10. RAM memory..... 16
- 3.1.11. Storage memory..... 16
- 3.1.12. Frequency..... 16
- 3.1.13. Audio Acquisition..... 16
- 3.2. Software functional requirements..... 16
  - 3.2.1. User interface Adaptability..... 16
  - 3.2.2. Offline mode..... 16
  - 3.2.3. Computing resources access..... 16
  - 3.2.4. Sensors Access..... 16
  - 3.2.5. Start-up/Exit time..... 17
  - 3.2.6. Application Size..... 17
  - 3.2.7. Power Management..... 17
  - 3.2.8. Augmented Reality Video Pipe..... 17
  - 3.2.9. Synchronization of AR Video Pipe and Rendering Pipe..... 17
  - 3.2.10. Replay Mode (Optional)..... 17
  - 3.2.11. Exposition of Camera and ISP statistics..... 17
- 3.3. Software non-functional requirements..... 18
  - 3.3.1. Portability..... 18
  - 3.3.2. Stability..... 18
  - 3.3.3. Extensibility..... 18
  - 3.3.4. Restricted Access..... 18
  - 3.3.5. Observability..... 18
  - 3.3.6. Scalability..... 18
  - 3.3.7. Debug..... 18
- References..... 19

## Executive Summary

### Scope

The original scope of this deliverable was to describe the final use-cases for the project and highlight the requirements needed on the next generation platform in order to achieve the implementation of such use-cases. As the project progressed, the VENTURI members decided to split the use-cases into several steps in order to gradually implement the various components with increasing complexity, both in terms of computation and in terms of integration.

During the Grenoble F2F, <https://VENTURI.fbk.eu/reserved/meetings/fourth-technical-meeting-grenoble/>, it was agreed, amongst the partners, that D2.1.2 would focus on the use-case for the second year only.

The system requirements for the second year use-case have already been analyzed in D2.2.2, taking into account:

- The requirements (generally expressed, independently from the second year use-case) from D4.2
- The preliminary ideas (that were still not finalized yet) on second year use-case.

This document therefore contains the final description of the second year use-case (chapter 2).

Since the preliminary ideas on the use-case for the second year, that were used to derive the associated system requirements, were still close to the final use-case, Section 3 of this document mainly mirrors D2.2.2 requirements. Please note that these requirements are “cumulative”, i.e. they were issued by both the first and the second year use-cases and some might not apply to the second year use-case (e.g. graphics).

### Audience

This deliverable is public.

### Summary

The use-case for the second year is described as a story line, presenting the sequence of environments and situations encounter by a visually impaired character in order to reach his goal, in our case buy a gift for his nephew from a shop in a shopping mall. The various parts of the storyboard are divided into phases that correspond to a set of algorithms for the geo-localization and guidance of the person. Each phase is characterized by a set of components implemented by VENTURI partners and an application that uses them for the given use-case phase.

The implementation and integration of the various components have been enhanced by specific characteristic of the current VeDi device and will be recommended for the next generation device. These requirements are summarised in this deliverable.

### Structure

This deliverable is composed of two main parts. Section 2 contains the description of the use-case for the second year and the applications that will support the use-case. Section 3 contains the short list of requirements for the use-case. A full list of these requirements is already available in [1].

## 1. Introduction

This document describes the use-case discussed, reviewed and approved by the consortium for the demonstrator to be implemented for the second year of the project. An extra document will be released to describe the demonstrator for the third year. The distinction between the two different demonstrators wasn't

planned at the beginning of the project, but has been introduced to enable the demonstration and evaluation of the various AR algorithms incrementally before the final demonstrator.

This deliverable also collects the requirements for a next generation platform in order to support the specific use-cases and algorithms used.

## 2. Use-case description: Personal Assistant for the Visually Impaired

### 2.1. Introduction

The VEDI-2.0 use-case demonstrator will focus on the scenario of an indoor Personal Assistant for the visually impaired. The rationale behind this use-case is founded upon the following reasons:

- The scenario of the indoor shopping mall environment loosens the strict constraints on lighting imposed in VeDi-1.0. Thanks to the on-going developments of vision algorithms, we are confident that we can address a less-controlled space.
- Reality augmentation is not only about visual overlays, but it should also explore other sensorial modalities. By restricting the modes of feedback offered to the user, i.e. non-visual means, the project will also focus on audio and haptic modalities.
- Indoor navigation and guidance is a challenging task, and the foreseen fusion of visual/audio scene analysis and sensor-based PDR will create an incremental platform for algorithmic and system evaluation.
- Through established links with several blind institutes in Italy, France and Germany, there is a real possibility that a VeDi 2.0 prototype technology could be created that would find a niche in the market and could enhance real people's lives.

### 2.2. Embodiment



In order to free the user's hands for a walking cane and world-interaction, the VeDi device will be mounted in a pouch dangling from the user's neck, possibly stabilized with a chest clip or chest band. This mounting modality also opens up the opportunity of using hand gestures in-front of the world-facing camera to interface with the system. NFC could also be used to understand if the cell phone is taken from the embodiment and moved around to understand the surroundings (scene classification, sign detection,...).



## 2.3. Storyboard

The following description is focused on Marc who suffers from retinitis pigments, meaning that his visual acuity has been reduced to 20/400 in his best eye.

Marc finds it difficult to find exactly what he wants in the shops nowadays. It's his nephew's 12th birthday tomorrow, and he would like to buy him his favourite toy, Lego.

He searches the Internet and finds a good price at the 'Super Shopping Mall' in the Toys 'R Us shop. From 'Google directions' he knows that Bus 73 will take him from his home to the Shopping Mall. He downloads all the maps necessary for the navigation outside and inside the shopping mall onto his VeDi device.

The route that Marc must take to achieve his goal has been decomposed into a list of Points:

- Point A: Marc's arrival at the bus stop
- Point B: just inside the Shopping Mall entrance
- Point B1: danger sign position
- Point B2: entering an open space inside the mall
- Point B3: exiting an open space inside the mall
- Point C: just outside the elevator, ready for entry
- Point D: exiting the elevator
- Point E: just outside the shop
- Point F: the Lego shelf

These Points and the path between them are described in the following paragraphs and summarized in Section 2.4, with some maps describing a possible scenario. The use-case will be described referring to it, but it is not limited to a prefixed scenario.

As an optional feature, the various applications that are covering the use-case could be automatically switched based on identified triggers. These phase triggers have been identified, but in the final implementation their feasibility is not yet guaranteed, so this functionality could be considered optional.

### 2.3.1. Phase-1: route from bus stop to shopping mall entrance

After taking the correct bus to the shopping mall Marc arrives at the bus stop nearest to the shopping mall (Point A). Once he steps off the bus, as his VeDi device knew his destination and had pre-downloaded the building plan from OpenStreetMap, VeDi guides him to the entrance.

Marc is aurally guided to the entrance of the mall (Point B), using the Pedestrian Dead Reckoning (PDR).

The geographical position of the entrance of the shopping mall could be used as the trigger for the next phase, equivalent to a geo-fence.

### 2.3.2. Phase-2: route inside the base floor of the shopping mall

The system switches modality to indoor navigation mode.

The VeDi device guides Marc from the entrance of the mall (point B) to the elevator that will bring Marc to the floor of the mall where the toy shop is located (point C).

In this use-case several checkpoints are foreseen. Alert signals are detected (point B1) and suggestions for the correct way to cross a potentially dangerous passage are provided. Crowded areas are detected and a re-route is offered in order to avoid these areas/corridors in favour of freer passages. Open areas are detected (point B2) and in this case an optional navigation algorithm, more suitable for open spaces is activated, based on Structure

from Motion algorithm. The end of the open area (point B3) is detected to switch back to the first navigation algorithm (PDR).

The use-case can be described as follows:

Marc senses there is a potential danger sign in front of him (point B1), he stops and asks the VEDI device for confirmation. The VEDI device visually detects the danger sign and provides an audio alert “walk slowly”. A lot of people are present in the mall’s main thoroughfare (point B3) and, as a result, VeDi re-plans a route for Marc that should be quieter, as it skirts the main central aisle in favour of a slightly longer peripheral route.

Due to the complexity of this phase it is split into three sub-phases: navigation and re-routing with the PDR, alert detection and optionally navigation in open spaces with SfM.

The trigger for the next phase could be the geographical position of the elevator. During the navigation on the lower floor of the mall, other triggers can be used to detect the various scenarios. These triggers are the coordinates of the entrance, location of an open space, the exit from an open space, and the detection of possible warning signs, by their bright colours.

### 2.3.3. Phase-3: Lift usage

In the preceding phase:

VeDi directs Marc to the closest lift (point C), as this is the safest path to the specific shop that VeDi believes Marc needs. VeDi guides him to the buttons using visual cues and then observes Marc's hand to make sure he presses the correct button. As Marc has had normal vision for most of his life, he has a good mental picture of the environment, but the re-assurances provided by VeDi makes his trip to the mall less stressful. The target floor in this scenario is the second floor.

The lift doesn't have aural feedback but VeDi sees the floor number 2 and also senses that the barometric pressure change corresponds to going up 2 floors, so Marc gets out of the lift on the correct floor.

The trigger in this case could be the lift door opening at the correct floor, reaching the point D.

### 2.3.4. Phase-4: second navigation in the mall

Marc is safely guided to the correct shop (from point D to point E) where he can buy his gift. VeDi knows the interior map of the shop and can therefore assist Marc close to the shelves where the Lego toys used to be.

For this phase a simplified navigation system, with respect to phase-2, is implemented since this phase does not introduce any further components.

### 2.3.5. Phase-5: navigation inside the shop

Marc is guided from the entrance of the shop to the shelves where the Lego toys used to be.

### 2.3.6. Phase-6: navigation on the target shelf

Once in front of the toy shelf, VeDi sees that there are several boxes close together and the platform uses the embedded camera to scan the shelf and detects candidate positions where the desired toy might be.

Two possible target candidates are detected, one in the lower part of the shelf and the second in a central position on the higher part of the shelf. In both cases, the visual signatures are considered to be compatible with the goal.

Like in the elevator, Marc's hand is guided towards the first candidate box. After grabbing it, Marc orients it in front of VeDi until the desired set-number is searched for. This procedure is repeated until the desired box is in Marc's hand.

### 2.4. Example Maps

The following pictures depict an example environment where the use-case can be demonstrated.

An example of the various phases has been sketched in the following diagrams. The areas depicted are based on the FBK office site, where the final demo will be executed. However, the algorithms being create can be applied to any scenario.

The path suggested by the VeDi device is represented by the red dotted arrow. Phase-1 is depicted in the following diagram:

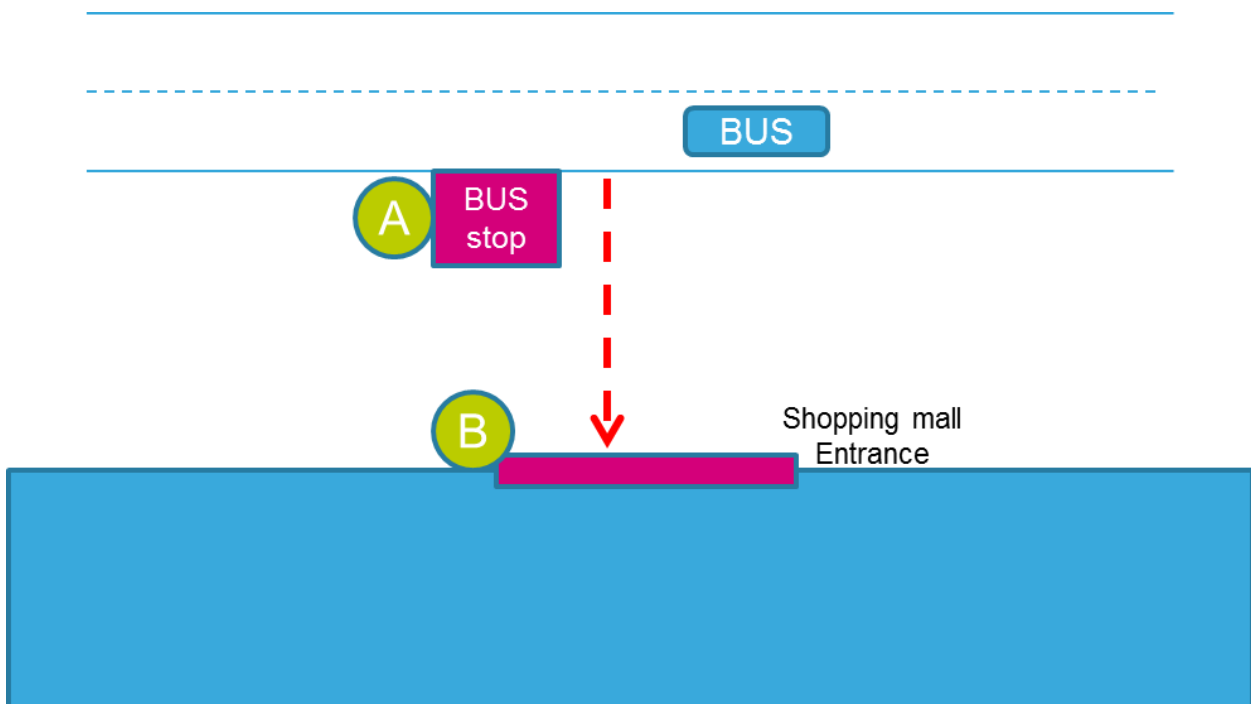


FIGURE 1: PHASE 1 - OUTDOOR NAVIGATION FROM BUS STOP TO SHOPPING MALL ENTRANCE

The next phase is the most complex one, since it involves several stages based on the types of areas crossed, on the warning sign encountered and on the possibility of having crowded areas.

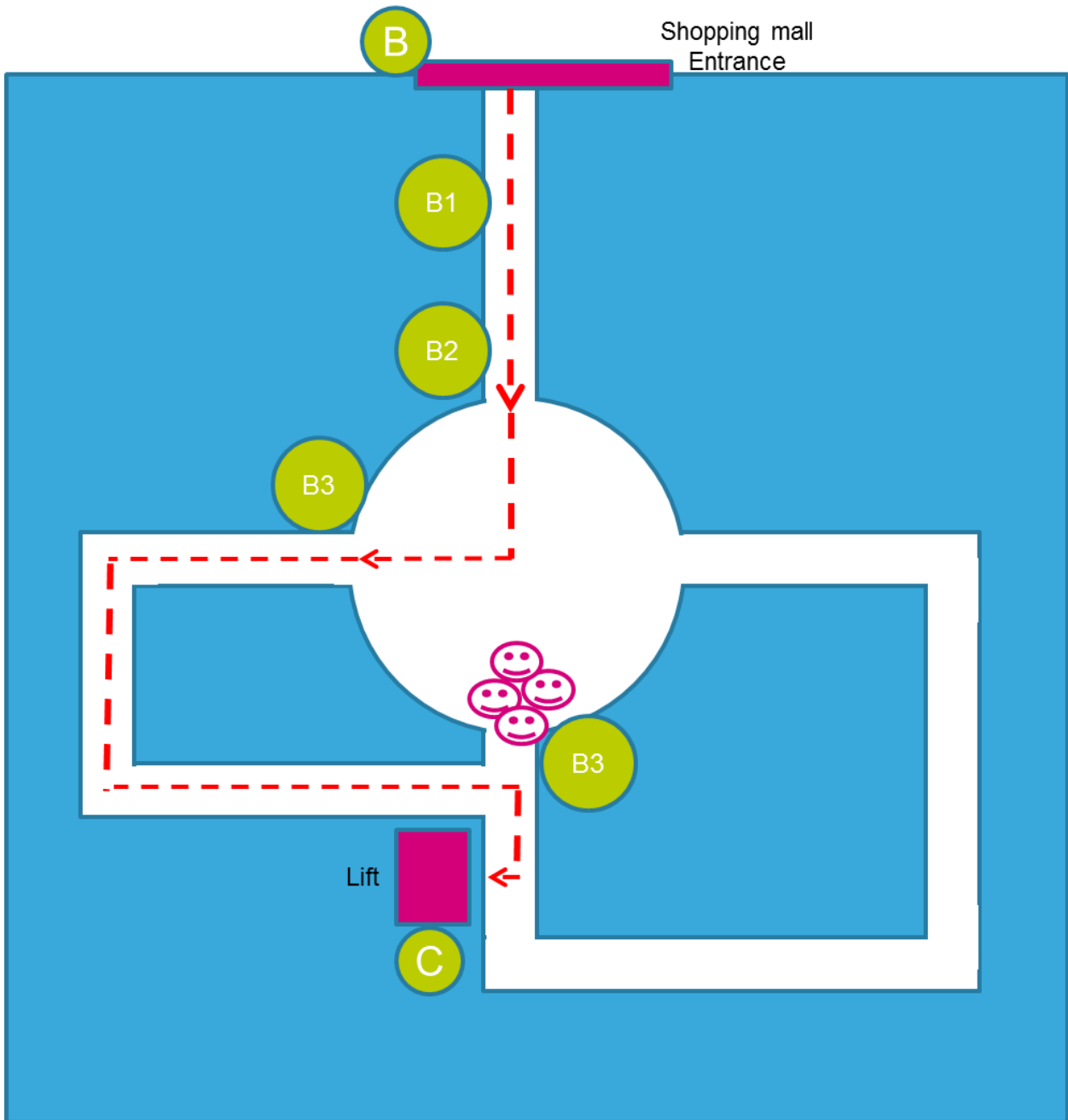


FIGURE 2: PHASE-2 – GUIDE TO THE LIFT AT THE BASE FLOOR OF THE SHOPPING MALL

The third phase, just outside the elevator and then inside it, is not depicted in a diagram since it is not a navigation application.

The fourth phase, conceptually analogous to the second one, is simpler since it includes only PDR navigation. It is depicted in the following figure:

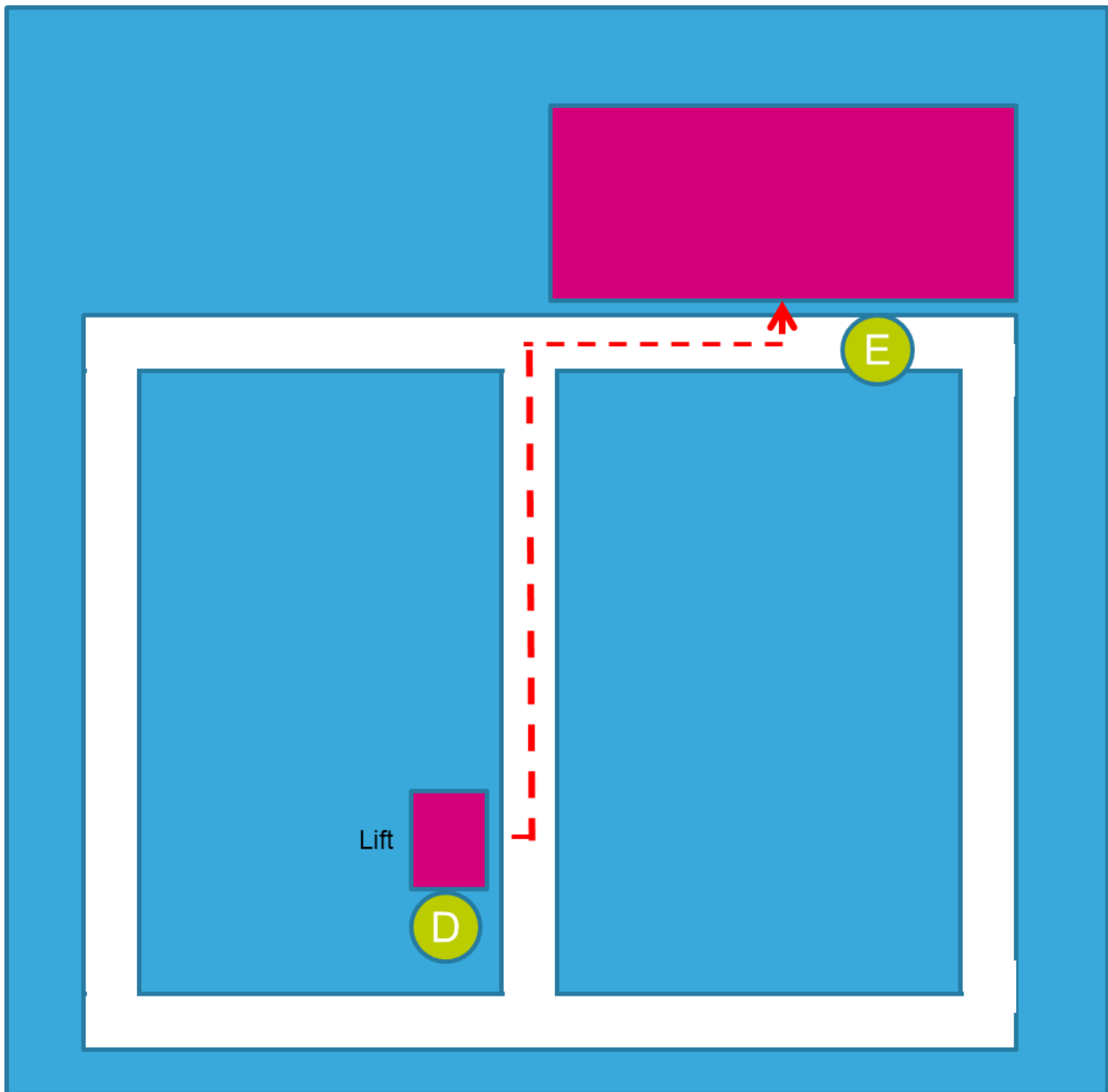


FIGURE 3: PHASE-4 – NAVIGATION AT THE FLOOR WHERE THE TARGET SHOP IS LOCATED

The next phase is the navigation inside the toy shop to the desired shelf.

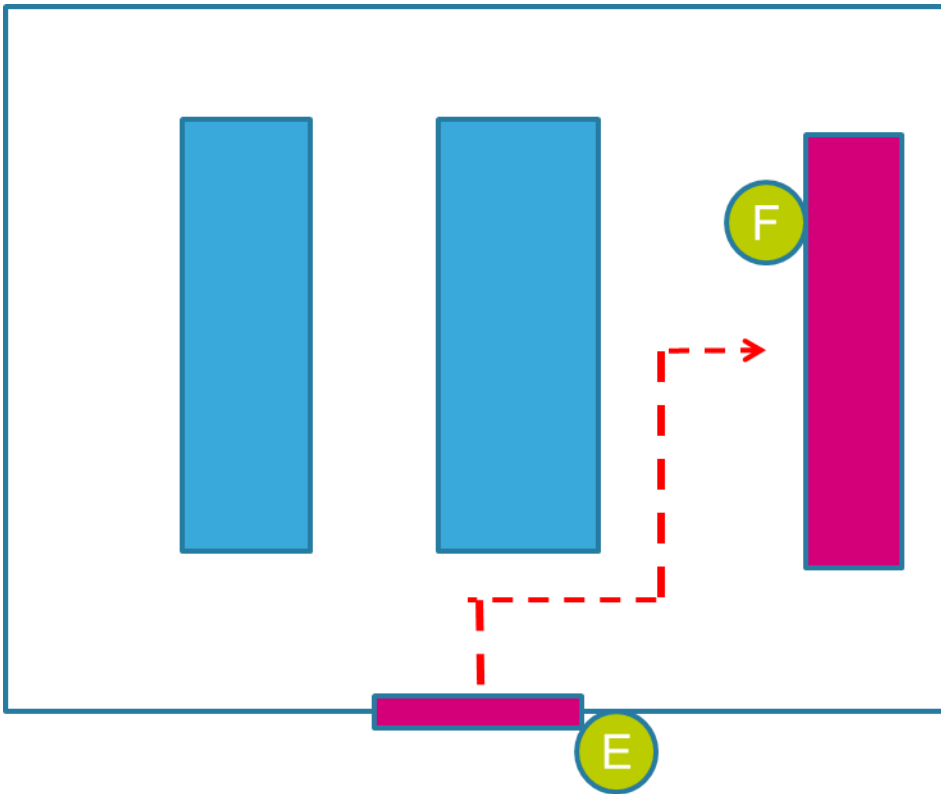


FIGURE 4: PHASE-5 – NAVIGATION INSIDE THE TOYS SHOP

The last phase is represented by the inspection of the LEGO shelf, the location of the good candidates for the target toy and the aural guidance to the box candidates to select the right one. Also in this phase a map is not provided as it is not a navigation problem.

### 3. Requirements for the actual device and the future devices

In deliverable D2.2.2 [1] all of the requirements for the use-case are listed and considered. Many of them are already fulfilled by the actual platform, but some of them can only be implemented on the next generation platform.

The requirements are divided into hardware functional requirements, software functional requirements and software non-functional requirements.

#### 3.1. Hardware functional requirements

##### 3.1.1. Camera

The hardware platform must support at least one front and at least one rear colour camera. Ideally, there should be a front stereo pair and two rear cameras that can work together as a stereo pair or even independently. Front and rear camera/s should be accessible simultaneously. The following camera parameters must be at least supported:

- Resolution needed for the AR analysis part: 640x480, grey-scale
- The frequency needed for the AR analysis part: 15fps.
- Resolution needed for the AR rendering part: colour images are ideally similar to the display resolution
- Frequency needed for the AR rendering part: 30fps.
- Latency should be feasible for a live view: max 180ms.

Spectral Sensitivity Functions of all cameras must be known and shared with VENTURI partners.

It must be possible to acquire still pictures (with or without auto-focus) at a resolution of at least 5 MPixels.

Moreover, image timestamp generation is required, ideally using the same clock as the one used for the inertial sensors.

### 3.1.2. Connectivity

The platform must support at least one form of cellular network connectivity and one wireless LAN network connectivity. Supported bandwidth must be at least 4Mbps, with a round trip time of 1000ms as a maximum.

Ad-hoc connectivity (without a server infrastructure) to other devices (either Bluetooth or WiFi) would be required for collaborative/social projection scenarios.

NFCs could also be used (for sub-use-case triggering) and will be considered, depending on the final outcome of the second year-Use-case definition.

### 3.1.3. Sensors

The platform must support the following sensors:

- 3 axis accelerometer,
- 3 axis magnetometer,
- 3 axis gyroscope,
- a barometric pressure sensor with a 3 meters relative precision (one level inside a building)
- a proximity sensor (to trigger sub-use-cases)

Sensor sampling frequency should be higher than the camera frame rate, and no less than 60 Hz. Sensors must provide at least 10-bit resolution samples and support range selection capabilities.

### 3.1.4. Input method

The platform must support the touch-panel as an input device.

The platform should connect to the Sony watch (used as input to trigger sub-use-cases).

### 3.1.5. Display

The platform must support a display with at least the following characteristics:

- Size: 3.7";
- Resolution: WVGA (480 x 854);
- Frame rate: 30fps;
- DPI: 270.

In addition, the platform should be able to support different screen form factors with no or little board modifications.

The platform must support dual display output channels, one for the device and the other to feed content to a pico-projector (most probably via HDMI).

### 3.1.6. Graphics Hardware

The platform must include a graphics chip capable of performing complex 3D-graphics rendering with at least 15000 polygons/sec, as a minimum, for all of the objects in the scene.

### 3.1.7. Audio

The platform must support stereo audio playback via an external headset. The PCM latency must not exceed 300ms and the DAC frequency must not be less than 44100Hz.

### 3.1.8. Power

The platform must be able to sustain at least 1 hour continuous operation, battery operated. This is directly linked to the software functional requirement at paragraph 3.2.7.

### 3.1.9. Autonomous mode

Besides development and debug mode, the platform shall be ran in autonomous mode, that is battery operated, and no debug console (UART, JTAG, other).

### 3.1.10. RAM memory

The platform must be equipped with at least **1 GB** of Random Access Memory.

### 3.1.11. Storage memory

The platform must be equipped with at least 16 GB non-volatile memory.

### 3.1.12. Frequency

Host processor peak value frequency must be at least 1GHz.

### 3.1.13. Audio Acquisition

Microphone audio circuitry on the platform is designed for recording speech: 16-bit at 44.1kHz. Better performance could be required for the next VeDi platform, depending on the result of the experiments (see T4.2).

## 3.2. Software functional requirements

### 3.2.1. User interface Adaptability

The System User Interface must be able to adapt to different screen sizes and form-factors with limited or no modifications to the application, e.g. Tablet or Smartphone

### 3.2.2. Offline mode

The terminal must be able to operate without network connectivity, using cached data.

### 3.2.3. Computing resources access

All non-critical platform computational resources must be accessible to the application. For example by means of standard APIs such as: OpenGL|ES, OpenCL, Renderscript, OpenMAX.

### 3.2.4. Sensors Access

The first sub-requirement is that the application must be able to access the platform's positioning sensor resources provided by the hardware platform by means of Operating System or other standard APIs. Possibility to access low-level data (to apply previously computed correction offsets) is desirable.

The second sub requirement related to sensors access is that the video frames have to be time-stamped in an accurate manner. This sub-requirement overlaps the Augmented Reality Video pipe requirement and is extensively discussed in 3.2.8 analysis section.

The third sub requirement is that the positioning sensor samples have to be time-stamped in an accurate manner.



The last sub requirement is to allow positioning sensor samples to be synchronised at the application level with the video frames, the sensor samples need to be time-stamped with the same time base as the video frames, in order to avoid offset and drift.

### 3.2.5. Start-up/Exit time

The application must meet average user expectations for start-up time: applications must start in less than 15 seconds, with a notification that progress is on-going (e.g. UI with a clock or progress bar).

The application must close gracefully and release all platform resources used during its operation. A shut-down progress bar must be displayed at application closure.

### 3.2.6. Application Size

The application installer package should not be bigger than 20 MB. This includes all code and application resources (icons, background images, etc.) but does not include any multimedia content, terrain model, POIs database, 3D model.

### 3.2.7. Power Management

The application should be able to run without interruptions for at least one hour with no external power supply. This requirement is strictly related to HF8 and should be treated as a transversal hardware/software requirement.

### 3.2.8. Augmented Reality Video Pipe

The camera resolution needed for the rendering part: colour images and ideally similar to the display resolution. Camera frequency needed for the rendering part: 30fps. Time-stamps for the camera images (ideally using the same clock as the one used for the inertial sensors) must be supported.

### 3.2.9. Synchronization of AR Video Pipe and Rendering Pipe

Video pipeline should provide two synchronized image qualities: one grey-scale low-resolution for the vision part, and one colour high-resolution for the rendering part.

### 3.2.10. Replay Mode (Optional)

To guarantee platform benchmarking repeatability, a modality to record and playback events occurring during platform operation in a specific use-case mode is needed. This modality will be added only if time and resources will be available.

### 3.2.11. Exposition of Camera and ISP statistics

The application must be able to access camera and ISP statistics by means of Operating System or other APIs. The operating system shall be capable of delivering camera statistics to the application, in viewfinder mode. The required statistics, which shall be exploited for context sensing and interpretation within the WP4, are:

- Exposure time;
- Aperture;
- Gain (ISO);
- White balance and possibly the mean values of the R, G, B colour planes;
- Focus;
- Global histogram (if available);
- Face position and size (if available);
- DCT coefficients (if accessible and available in viewfinder mode).

All this information can help computer vision techniques, based on image analysis, in automatically identifying the scene context within VeDi-2 and VeDi-3 demonstrators. In fact, all the aforementioned parameters strongly depend on the capturing conditions (e.g. day/night, indoor/outdoor, landscape/close-up, etc.).

### **3.3. Software non-functional requirements**

#### **3.3.1. Portability**

The application will run on Android OS. Nevertheless, applications shall be architected in a way to ease portability among different mobile operating systems (C/C++ code that can be recompiled on different OS).

#### **3.3.2. Stability**

The application shall not exhibit force closes or hangs. The performance of other applications running on the platform shall not be impaired by VENTURI stack.

#### **3.3.3. Extensibility**

The application shall be extendable with new functionalities without requiring user's manual intervention.

#### **3.3.4. Restricted Access**

Application shall enforce basic access control for privacy-sensitive contents: personalized access or secure login capabilities.

#### **3.3.5. Observability**

The VENTURI system shall be observable by means of software or hardware assisted profiling and tracing tools for performance and power consumption characterization.

#### **3.3.6. Scalability**

The VENTURI system shall be able to scale well with respect to the number of users and the number of augmented reality content providers.

#### **3.3.7. Debug**

The Software platform shall expose software ports for application and operating system debugging purposes, possibly using de-facto standard tools.

## References

[1] D2.2.2 “Refined Detailed Design Specifications for STE L9540-based Platform”.