ROBOTS IN ASSISTED
LIVING ENVIRONMENTS

UNOBTRUSIVE, EFFICIENT, RELIABLE AND MODULAR SOLUTIONS FOR INDEPENDENT AGEING

Research Innovation Action

Project Number: 643892 Start Date of Project: 01/04/2015 Duration: 36 months

DELIVERABLE 3.3

Conceptual architecture for sensing methods and sensor data sharing III

Dissemination Level	**Public**
Due Date of Deliverable	Project Month 24, March 2017
Actual Submission Date	24 April 2017
Work Package	WP3, *Modular conceptual home architecture design and ICT method development for efficient, robust and flexible elder monitoring and caring*
Task	T3.1, *Conceptual architecture for sensing methods and sensor data sharing*
Lead Beneficiary	S&C
Contributing beneficiaries	NCSR-D, TWG
Type	Report
Status	Submitted
Version	Final

Project funded by the European Union's Horizon 2020 Research and Innovation Actions

Abstract

Architecture document, pertaining to the information dependencies and interoperability between the ADL recognition components. This document documents the final architecture at the end of WP3.

History and Contributors

Ver	Date	Description	Contributors
01	12 Oct 2016	First draft, assuming D3.2 as a starting point.	S&C, NCSR-D
02	14 Dec 2016	List of audio-visual analysis components planned for deployment at FHAG and FZ trials (Section 3).	NCSR-D
03	20 Dec 2016	Revisited list of clinical requirements and relevant technologies (Section 2), taking into account the addition of Instrumental ADLs (D2.3, Oct 2016) and technical developments (Bochum plenary, Dec 2016).	NCSR-D
04	6 Mar 2016	Finalized list of system components to be deployed at FHAG and FZ trials (Section 2).	NCSR-D, TWG, S&C
05	15 Mar 2017	Specification of the nature of the interfaces between components. Updated architecture diagram with new components in D3.5 (Section 5).	NCSR-D
06	16 Mar 2017	Specification of the smart home sensors used to recognize ADLs by the EnControl rule engine (Section 2).	S&C
07	19 Apr 2017	Revised the list of clinical requirements (Section 2) based on D2.7 and the acoustic recognition of mood (Section 3).	NCSR-D
08	24 Apr 2017	Updated architecture diagram and the interfaces between components (Section 5) to reflect the finally submitted integrated system D3.9/D3.10.	NCSR-D, TWG, S&C
09	24 Apr 2017	Internal review	TWG
Fin	24 Apr 2017	Addressing internal review comments, final document preparation and submission	NCSR-D

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Purpose and Scope

This deliverable is the final version of the *conceptual architecture* that specifies the information dependencies and interoperability between RADIO components for satisfying the ADL recognition requirements set in WP2.

Within the scope of this document is:

- To identify the recognition methods that will be extended and adapted in the context of Task 3.2 in order to extract the interRAI assessment items that are required by WP2.
- By studying the information dependencies between these components, to specify the conceptual architecture of the RADIO Home system that integrates the processing elements executing the recognition components above with the sensing elements (including smart home and robot-mounted sensors).

Outside the scope of this document is:

- To specify the physical architecture, such as whether components will execute on FPGA boards or general-purpose computers. This is dealt with in Task 4.1.
- To specify the architecture (either conceptual or physical) of the communication between the RADIO Home and other nodes of the RADIO ecosystem, such as cloud storage components and components meant to be used by hospital personnel or informal care-givers. This is dealt with in Task 5.1.

1.2 Approach

This deliverable documents work in Task 3.1, which analyses the user requirements collected by work in WP2 in order to develop the system architecture. The work plan in this task includes:

- Identifying recognition methods that can cover requirements on collecting information.
- Studying the data interdependencies between the different recognition methods in order to specify what intermediate results can be shared. This leads to providing technical requirements regarding data processing and data transfer.
- The analysis above can also identify alternatives (data and processing) for extracting the same information item. This can be used to balance between robustness and ethical requirements. For example, suppose bathroom ADLs can be recognized using only acoustic data or (more robustly) using a combination of acoustic and visual data. This deliverable will describe the alternatives in order to provide input to the relevant ethical discussion in WP2.

S&C leads this task and provides knowledge and experience with current smart home systems to match requirements from WP2, and the information exchange between the smart home system and the robot. **NCSR-D**, **TWG**, and **AVN** carry out the analysis that identifies appropriate recognition methods for the interRAI assessment items. **TWG** and **RUB** explore different approaches focusing on WSN and short range wireless communication technologies that can be exploited in RADIO.

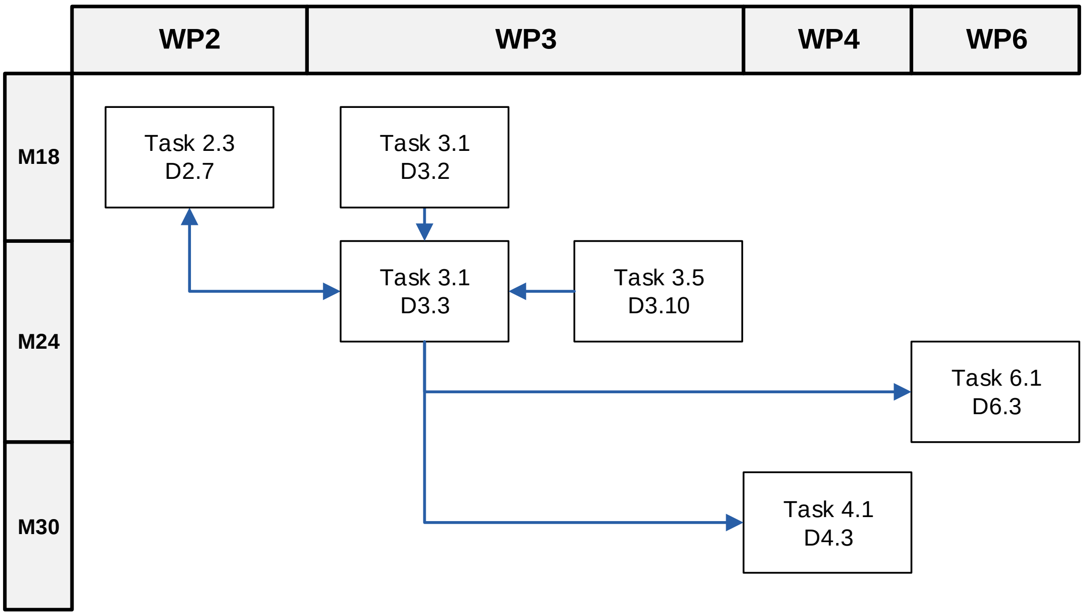

Figure 1: Relation to other Work Packages and Deliverables

1.3 Relation to Other Work Packages and Deliverables

This document is part of a cluster of closely related deliverables. The initial conceptual architecture (D3.1) planned development work until M18, including a first overview of the state of the art upon which RADIO development was based. The next version (D3.2) had the architectural design and component interconnections needed to satisfy the user requirements (WP2) and the findings of D3.6 *Network robustness and efficiency methods*.

In this manner, work around ADL and Mood recognition is organized as follows:

- D3.2 *Conceptual architecture II* (R): design and interconnections of components
- D3.4 *ADL and mood recognition methods I* (R): methods developed in order to implement the components specified in D3.2
- D3.9 *Integrated data analysis system I* (DEM): prototypes of the methods from D3.4

This final version records adjustments necessitated as WP2 and WP3 progressed, but also bi-directionally influences the development of requirements in T2.3 and prototypes in T3.5. This final version documents the architecture and interfacing to include the final WP3 methods (D3.5, D3.7, D3.8), as prototyped in D3.10.

The first version of the architecture (D3.1) was used in WP2 to drive the discussion on balancing between clinical requirements, obtrusiveness, and technical feasibility. The outcome of this discussion was the *Guidelines for balancing between medical requirements and obtrusiveness* (D2.6 and D2.7) is used by this document to document where the different recognition items stand with respect to obtrusiveness.

Finally, this deliverable will be used by Task 4.1 in order to prepare the third version of the physical architecture, D4.3 *Architecture for extending smart homes with robotic platforms*.

2 RECOGNITION OF INTERRAI ITEMS

We assume as a starting point the interRAI assessment items that have been marked as relevant to RADIO in D2.2/D2.3. Specifically, we assume the items marked as "Observation", since these are assessed by observing the user. Other interRAI items are assessed by interviewing the user or by accessing the medical record, and are outside the scope of work in WP3.

For each of the observation items, Tables 1, 2 and 3 indicate whether they have been treated in WP3, and, if so, the kind of data and processing over these data that is used to assess the item. Specifically, we indicate in the *Status* column of Tables 1, 2 and 3:

1. Item where it is unclear how this recognition can be achieved, as it involves extracting qualitative information that is beyond the state of the art and would need novel algorithmic design. The only viable means is providing raw audio-visual content for observation by human caregivers, should this be acceptable based on the discussion in D2.6 and D2.7.
2. Items that can be recognized using RADIO algorithms and methods, but at an accuracy that compromises clinical assessment results and functional unobtrusiveness.
3. Items that can be recognized using RADIO algorithms and methods at an acceptable accuracy and are included in the RADIO architecture.

Table 1: ADL-related data extraction methods

Item	Name	Technology	Status	Comments
ADL Self Performance	Bathing, personal hygiene, Dressing upper body; dressing lower body	Audio-visual event recognition	1	Extracting qualitative information beyond the state of the art.
	Walking, locomotion	Human pattern recognition (HPR) in 3D or range data and face identification.	3	Assuming the scene is not crowded, or under fusion with some user identification method external to HPR. Choice between pure range data or fused range/visual/3D data based on obtrusiveness.
	Toilet transfer	Visual event recognition. Proximity sensor.	1	Automatically extracting qualitative information without human observation.
	Toilet use	Audio event recognition	3	Event log only, without qualitative information
	Bed mobility	Visual event recognition and smart home sensing	3	Measuring the time it takes to transfer out of bed and armchairs.
	Eating	Visual event recognition	1	Automatically extracting qualitative information without human observation.
Locomotion and Walking	Timed 4-meter walk; distance walked	Human pattern recognition (HPR) in 3D or range data and face identification.	3	Assuming the scene is not crowded, or under fusion with some user identification method external to HPR. Choice between pure range data or fused range/visual/3D data based on obtrusiveness.
	Distance wheeled self : *N/A, as ability to walk is required (cf. inclusion criteria, Sect 2, D2.2)*			
Activity Level	Total hours of exercise or physical activity in last 3 days	Human pattern recognition (HPR) in 3D or range data and face identification.	3	Assuming the scene is not crowded, or under fusion with some user identification method external to HPR. Choice between pure range data or fused range/visual/3D data based on obtrusiveness.
	Number of days went out of the house in last 3 days	Smart home sensing	3	Open/close sensor for main door. Detecting exit and entrance.

Table continues from previous page.

Item	Name	Technology	Status	Comments
Activity Preferences and Involvement	Cards, games, or puzzles; crafts or arts; reading, writing, or crossword puzzles	Audio-visual event recognition	1	Although it might be possible to make some location-based inferences, such inferences cannot estimate the ability to meaningfully carry out these activities.
	Computer activity	Audio-visual event recognition	1	Although it might be possible to make some location-based inferences, such inferences cannot estimate the ability to meaningfully carry out these activities.
		Computer activity logging	1	Might be possible, but would be a new line of research outside the scope of RADIO
	Conversing or talking on the phone	Audio-visual event recognition	1	Speaker identification and visual recognition of device usage can be used to recognize talking/holding phone set (at a smaller degree of accuracy) but cannot assume it is a meaningful conversation.
	Dancing, exercise or sports	Audio-visual and range data event recognition	2	Possible, at some degree, by fusing and training the existing RADIO methods and smart home sensors.
	Gardening or plants: *N/A, RADIO operates indoors*			
	Helping others	Audio-visual event recognition	1	Automatically extracting qualitative information without human observation.
	Music or singing	Audio event recognition	3	Audio event classification, relatively robust.
	Pets: *N/A, presence of pets is not considered in the RADIO experiment environment*			
	Spiritual or religious activities	Visual event recognition, automatic speech recognition	1	Extracting qualitative information and language understanding beyond the state of the art.
	Trips/shopping: *N/A, RADIO operates indoors*			
	Walking or wheeling outdoors: *N/A, RADIO operates indoors*			
	Watching TV or listening to radio	Fused smart home and audio-visual event recognition	3	Detecting TV energy consumption, detecting if the right user is sitting in chair, on the couch and in bed and detecting if the chair is closer to the TV location.
	Discussing and reminiscing about life	Automatic speech recognition	1	Extracting qualitative information and language understanding beyond the state of the art.
Time Asleep During Day	Time asleep during day	Audio-visual event recognition	1	Extracting qualitative information beyond the state of the art.

Table 2: Cognitive and Mood Items data extraction methods

Item	Name	Technology	Status	Comments
Cognitive skills for daily decision making				
Memory/ recall ability				
Periodic Disordered Thinking Or Awareness			1	Extracting qualitative information and language understanding beyond the state of the art.
Acute Change in Mental Status from Person's Usual Functioning				
Indicators of possible depressed anxious or sad mood	Made negative statements	NLP in Facebook posts	3	Information extraction from posts.
	Persistent anger with self or others	NLP in Facebook posts	2	Challenging information extraction. State of the art not accurate enough to be usable.
	Expressions, including non-verbal, of what appear to be unrealistic fears	NLP, speech recognition, facial expression analysis	1	Extracting qualitative information and language understanding beyond the state of the art.
	Repetitive health complaints; repetitive anxious complaints and (non-health related) concerns	NLP in Facebook posts	2	Challenging information extraction. State of the art not accurate enough to be usable.
	Sad, pained, or worried facial expressions	Facial expression analysis		
	Crying, tearfulness	Audio event recognition		
	Recurrent statements that something terrible is about to happen	NLP in Facebook posts	2	Information extraction from posts.
	Withdrawal from activities of interest; Reduced social interactions	Facebook activity analysis. Speaker identification for visitors.	2	Log of interactions, frequency.
	Expressions, including non-verbal, of a lack of pleasure in life	NLP, speech recognition, facial expression analysis	2	Challenging information extraction. State of the art not accurate enough to be usable.
Behaviour Symptoms	Wandering		1	
	Verbal abuse		2	
	Physical abuse		2	Challenging extraction tasks. State of the art not accurate enough to be usable.
	Socially inappropriate or disruptive behaviour		1	
	Inappropriate public sexual behaviour or public disrobing; Resists care		1	

Table 3: Health Conditions and Continence items data extraction methods

Item	Name	Technology	Status	Comments
Falls		Visual and range pattern recognition.	3	Choice between pure range data or fused range/visual/3D data based on obtrusiveness.
Health Condition Frequency	Difficult or unable to move self to standing position unassisted; to turn self around and face the opposite direction when standing	Visual event recognition	3	Chair transfer measurement
	Dizziness	Visual event recognition	1	Extracting qualitative information beyond the state of the art.
	Unsteady gait	Human walking pattern recognition in range data.	2	Challenging pattern recognition, not accurate enough to be usable.
	Difficulty clearing airway secretion; Constipation; Diarrhoea	Audio event recognition	1	Extracting qualitative information beyond the state of the art.
	Vomiting	Audio event recognition	3	
	Difficulty falling asleep or staying asleep; waking up too early; restlessness; non-restful sleep; Too much sleep	Audio-visual event recognition	2	Challenging pattern recognition, not accurate enough to be usable.
	Aspiration, coughing while eating	Audio event recognition	2	Challenging pattern recognition, not accurate enough to be usable.
Dyspnea		Audio-visual recognition	1	Extracting qualitative information beyond the state of the art.
Fatigue		Audio-visual recognition	1	Extracting qualitative information beyond the state of the art.
Pain symptoms	Frequency with which person complains or shows evidence of pain (including grimacing, teeth clenching, moaning, withdrawal when touched, or other nonverbal signs suggesting pain) Consistency of pain		2	Challenging audio-visual event recognition, not accurate enough to be usable.
Continence	Bladder continence		1	

Table 4: Instrumental ADLs

Item	Name	Technology	Status	Comments
ADL SELF-PERFORMANCE AND CAPACITY	Meal preparation—How meals are prepared (e.g., planning meals, assembling ingredients, cooking, setting out food and utensils)	Smart home sensing and audio visual event recognition	3	Recognized by fusing smart home data such as the time of day, the user's movement, the on/off status of various electrical appliances such as the microwave, the kettle, the toaster, and the cooktop and the open/close detection of the fridge and cupboards.
	Ordinary housework—How ordinary work around the house is performed (e.g., doing dishes, dusting, making bed, tidying up, laundry)	Audio event recognition	2	
	Managing finances—How bills are paid, checkbook is balanced, household expenses are budgeted, credit card account is monitored	Audio-visual event recognition	1	Extracting qualitative information and language understanding beyond the state of the art.
	Managing medications—How medications are managed (e.g., remembering to take medicines, opening bottles, taking correct drug dosages, giving injections, applying ointments)	Visual event recognition	3	A first method of pill intake has already been tested during the intermediate phase pilot studies.
	Phone use—How telephone calls are made or received (with assistive devices such as large numbers on telephone, amplification as needed)	Visual /depth pattern recognition	1	Extracting qualitative information beyond the state of the art.
	Stairs—How full flight of stairs is managed (12–14 stairs) *N/A, stairs are out of reach*			
	Shopping—How shopping is performed for food and household items (e.g., selecting items, paying money): *N/A, RADIO operates indoors*			
	Transportation—How travels by public transportation (navigating system, paying fare) or driving self (including getting out of house, into and out of vehicles): *N/A, RADIO operates indoors*			

3 RECOGNIZING ADLS AND MOOD IN AUDIOVISUAL DATA

Table 5 below gives the current list of ADL recognition components developed in WP3 and other components relevant to ADL recognition, logging, and reporting.

Mood recognition has been restricted to processing Facebook posts using the RADIO RadFa app (D3.8). The state of the art in audio-visual mood detection has an accuracy that compromises clinical assessment results and functional unobtrusiveness. For example, Mano et al.[1] discuss the use of patient images and emotional detection to assist patients and elderly people within an in-home healthcare context. The proposed method achieves an overall accuracy of around 60% for 6 basic emotional classes, namely: neutral state, joy, sadness, fear, anger and surprise. Han et al.[2] propose a novel approach based on *deep neural networks (DNN)* that estimate the emotional probability distribution for speech segments. Even though experimental results demonstrate to accuracy improvement compared to the previous state of the art, the reported accuracy is around 50% for speech segments categorized in five emotions, namely: excitement, frustration, happiness, neutral and surprise.

Table 5: ADL recognition components and their auxiliary components

Component	Description
HumanPatterRecognition	Recognizes human walking patterns in laser scans and tracks walking. See also D3.4, Section 2.
ROSVisual	Tracks moving objects in the RGB/depth modality [D3.4, Section 3] and classifies motion as bed or chair transfer. See also D3.5, Section 2.
Motion Analysis	Recognizes motion and classifies it as “bed transfer” and “pill intake” events. See also D3.4, Section 4.
AUROS	Recognizes talking, watching TV, listening to music, and doing housework by acoustic analysis of the audio modality. See also D3.5, Section 3.
Home automation sensors	Recognizes presence, manipulation of appliances (such as TV or kitchen appliances), and opening/closing of cupboards. See also D3.5, Section 4.
BLE Localization	A system of applications for BLE beacons, BLE gateway, and mobile device (including the robot) for localizing mobile devices using the beacon signal. See also D3.5, Section 5.
HPR Wrapper	Uses HumanPatterRecognition output to recognize and time “walked 4m” events.
Motion Analysis Wrapper	Uses the output from motion analysis to time the bed transfer event.
ROSVisual Wrapper	Uses the output from ROSVisual to time chair and bed transfer events and to recognize and time “walked 4m” events.

[1] Mano, Leandro Y. et al. "Exploiting IoT technologies for enhancing Health Smart Homes through patient identification and emotion recognition." Computer Communications 89 (2016): 178-190.
[2] Han, Kun, Dong Yu, and Ivan Tashev. "Speech emotion recognition using deep neural network and extreme learning machine." Interspeech. 2014.

Deep learning is also adopted by Mao[3] where a convolutional neural network is used to estimate the emotion of speech segments on four datasets, reporting accuracy rates of around 80%. As the author note, however, this result is attained using *professional actors* as subjects, *deliberately pronouncing particular sentences*, in a *controlled acoustic environment*.

In general, state-of-the-art emotion recognition in controlled environments but using non-professional subjects and not deliberately pronounced emotions achieves overall performance of around 50% for only five classes, even in controlled environments that are relatively distant to a real home monitoring context.[4] Finally, the results of the realistic *emotion recognition in the wild* challenge are worth mentioning, where the winning system applied an audio-visual fusion approach where emotion video clips are regarded as an image set and different kinds of image set models are used to represent the video clips as a collection of data points on Riemannian manifold. A score-level fusion of classifiers learned based on different kernel methods and different modalities is conducted for final recognition results. Evaluation on the EmotiW2014 (7 classes: angry, disgust, fear, happy, neutral, sad, and surprise) benchmark has demonstrated the best performance was around 50%.[5]

[3] Mao, Qirong, et al. "Learning salient features for speech emotion recognition using convolutional neural networks." IEEE Transactions on Multimedia 16.8 (2014): 2203-2213.

[4] Martin, Olivier, et al. "The enterface'05 audio-visual emotion database." Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on. IEEE, 2006.

[5] Liu, Mengyi, et al. "Combining multiple kernel methods on Riemannian manifold for emotion recognition in the wild." Proceedings of the 16th International Conference on Multimodal Interaction. ACM, 2014.

4 RECOGNIZING ADLS IN SMART HOME SENSOR DATA

4.1 Introduction

The smart home architecture in RADIO has several requirements that need to be fulfilled. The product that RADIO will employ for creating the smart home solution is called *enControl*™. *enControl*™ provides an initial solution for home automation, and will be enhanced and upgraded complementing other technologies and developments within RADIO that will form the final RADIO solution.

enControl™ is a white label B2B product by Sensing & Control Systems providing a smart home solution.

At a glance, *enControl*™ has four main functions:

1) Comfort
2) Security
3) Energy Management
4) Automation

Users of *enControl*™ are able to monitor and control remotely their homes & business using smartphones or any device running Web browsers, like tablets or PCs.

enControl™ can be divided into three main components, (i) home devices, (ii) IoT Platform and (iii) graphical user interfaces (as shown in Figure 2).

At home level, *enControl*™ is composed by (i) end devices (sensors and actuators) and (ii) home controller (Gateway). The gateway supports communication using different protocols, but mostly uses ZWAVE radio technology for the communication of off-the-shelf end devices and home controller.

There are more than 300+ companies providing ZWAVE with more than 1300 products[6] which enables *enControl*™ to deliver the four main functionalities highlighted above. The following list[7] extends the four main functionalities from the point of view of the information triggered by the end devices:

1) Comfort
    a. Climate monitoring
    b. Climate control
    c. Temperature, Humidity, CO_2 (etc…) levels
2) Security
    a. Detection of door/window opening
    b. Detection of movement
    c. Detection of Smoke
    d. Detection of CO
    e. Detection of water basement
3) Energy Management
    a. Energy consumption
    b. Energy control (switch on/off electricity, water, gas, etc…)
4) Automation
    a. Switch on/off appliances
    b. Switch on/off lights
    c. Open/Close doors, curtains, shutters

[6] According to http://z-wavealliance.org

[7] The list does not pretend to cover all possibilities, the Reader should understand that the smart home solution can integrate any ZWAVE standard product, thus enabling the functionality delivered by a particular product. For a full list of products, please visit ZWAVE alliance web page.

By exploiting or aggregating smart home information per se, and/or with other RADIO components information, insights on the activity being generated by sensors can provide means of inferring the end-users' activity. The combination of the smart home with other RADIO components provides an assisted living environment that does not rely on any intrusive devices, such as wearables attached to the end-user.

The home controller is a product from Sensing & Control, and complements the smart home functions delivered by the IoT platform.

The IoT platform contains the core of the smart home solution. It provides an open REST API enabling the home controllers to exchange information bi-directionally, based on synchronous or asynchronous actions triggered by IoT or end users through *enControl*™ interfaces.

The IoT platform serves as the main repository of information, keeps historical information about end devices data (information and status) as well as basic actions triggered by different actors, so users of the smart home solution know in real time who-when-what of actions monitored (for instance "the TV set has been switched on by Maria on 11 July 2015, 20:30 CET").

The open REST API is used by user interfaces through web clients and smartphone apps in order to present to end users the information being acquired from home end devices, and the action than can be triggered to them, so it encapsulates the offered smart home functions.

It is important to notice that *enControl*™ clients (smart phone apps and Web interfaces, 3rd party services, etc.) require internet connection to interact with the smart home through the API. It should be noted that experience with deployment of the solution is that this introduces no perceptible delay in executing actions by ZWAVE devices.

The main home controller target is to timely exchange information between devices and the IoT Platform (and thus with end users), and information being generated by the home controller itself. However, it additionally provides three important functions:

1) Management of ZWAVE network and devices
2) Basic pre-processing capabilities (mainly energy calculations)
3) Local information repository in order to deal with temporal internet cut offs (so historical information is kept to be sent when internet connectivity is resumed).

The current smart home solution provides an intuitive and user-friendly interface, which is similar for tablets, PCs (web interface) and smart phones.

4.2 Home Automation Conceptual Architecture

Figure 2 represents *enControl's*™ service architecture. Following up on the high level details about overall service provided above, this section focuses on those aspects of *enControl*™ that facilitate the integration of the current *enControl*™ version into the RADIO's AAL solution.

The IoT platform is able to connect, transport, process, analyse and manage data from the sensors to the real world to high-level applications and vice versa. It is able to operate in a booth wireless/wired network environments and supports different communication protocols. As a function summary:

- **Collect data**

  Collect data refers to the ability to retrieve and store information for further exploitation about all the parameters that are relevant to the system: messages, device status, commands, errors, exceptions.

- **Analyse data**

  The IoT Platform has processes in charge of analysing data and taking decisions based on the quality/importance/integrity of this data.

- **Data Aggregation/Data Fusion**

  Data aggregation refers to the ability to concatenate info from devices, for example when there is a mix of information between data from mash-up sensors.

  Data fusion, is the process of integration of multiple data and knowledge representing the same real-world object into a consistent, accurate, and useful representation.

- **Translate data**

  This feature is related to the ability to transform raw data in bytes from the devices to high-level information.

- **System monitor**

  The IoT Platform is able to monitor the communication between devices and high level apps; and the overall status of the platform and device network. For instance, it is used to monitor key performance indicators like the number of messages sent per minute, number of exceptions/errors, number of devices connected to the platform, etc.

- **Transfer**

  The platform provides mechanisms to publish subscribe information to the queues exposed by the SDK.

- **Audit information**

  In order to enable traceability of the functionality of all sub-systems, active logging is implemented within the software components. Log files are stored and made available for analysis.

- **Secure components and communications**

  Communication between components and devices are secured.

The current solution of *enControl*™ provides means of interaction with home devices by accessing dedicated functions of the open API of the IoT platform. This implies that other ICT solutions (either at home like other RADIO components or remotely) willing to interact with *enControl*™ must have internet connection.

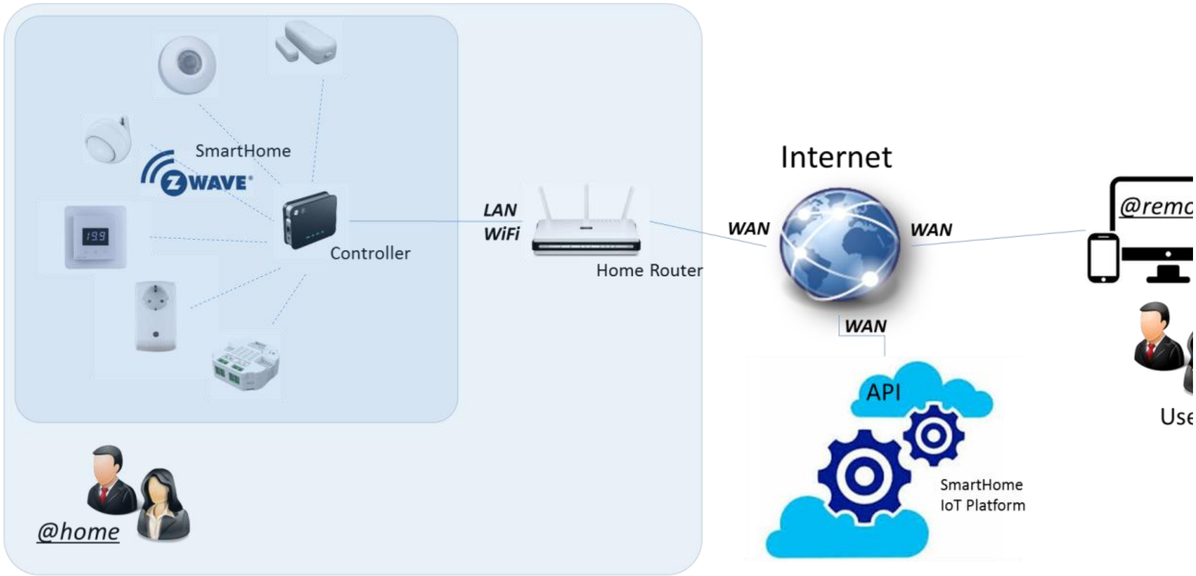

Figure 2: enControl™ *Service Architecture*

The API is divided into following main groups:

1) Authorisation and Authentication
    a. Functions used to grant access to different IoT resources and API functions
2) Devices
    a. Functions related to push and retrieve device status and data (including historical data)
    b. Functions related to the monitoring of ADLs
3) Users
    a. Functions related to the management of users
4) Installations
    a. Functions related to the management of smart home installations
5) Monitor
    a. Functions related to the status monitoring of the IoT

On the other hand, the smart home controller has computational resources as well as physical interfaces that can be used to connect third-party transceivers so to enable connectivity via WiFi or Bluetooth for example. The controller is a compendium of Java programs, from the Hardware Abstraction Layer (HAL), devoted to ZWAVE communication, and other available HW transceivers, to the smart home application (controller part) running on top of a Linux-enabled controller. For RADIO, *enControl*™ controller will be delivered running on Raspberry Pi.

Any additional code targeted to run on the home controller should only minimally affect the running main functions. Otherwise the responsiveness of the controller in front of potential user actions may be compromised leading to poor performance, which in turn may trigger potential rejection of the overall solution.

The smart home functionalities to be exploited for ADL recognition have already been listed in Sect 2.

5 RADIO CONCEPTUAL ARCHITECTURE

In this chapter, we combine the components described previously into the second version of the RADIO Conceptual Architecture. These components are:

1) The robotic platform
2) Wireless sensor networks
3) Smart Home
4) ADL and mood recognition methods

The Robot is an important element in the RADIO architecture concept. It is an entity that will continuously be in "motion & tracking" elder person activity. The robot is targeted to be the main interface with elderly people, but it is capable to hold computational resources and sensors needed by the recognition methods. Whether or not, the robot will hold them will be decided in future steps within WP4.

The goal of the smart home is to provide means of interaction with different controlled parts of a house, for instance open/close doors, curtains or shuts. Thus, it will provide home automation capabilities at a finger level of the elderly. The smart home thus will be a slave automation component in the RADIO architecture of the applications tailored in WP5. Furthermore, we will complement the data acquisition for the ADL and mood recognition methods, based on the huge amount of off-the-shelf sensors. Sections 2 and 3 provided an initial relation of InterRAI assessment items and RADIO Home data capture capabilities. This initial relation has been provided for the sake of future discussions, and will be updated in next iterations of this document.

All components need to be connected forming the RADIO communication system. RADIO will use common communication technologies between the main components (IP based communications, WIFI or LAN). Section 4 provided an initial overview of main important requirements for the connection of different sensors and actuators in the RADIO architecture, where low power wireless sensor technologies are the target.

Figure 3 depicts the architecture of the RADIO Home communication channels. This architecture does not reflect the physical wireless communication channels which are specified in WP4. It also does not reflect how multiple RADIO Homes, caregivers' devices, and care services interconnect into the RADIO Ecosystem, developed in WP5.

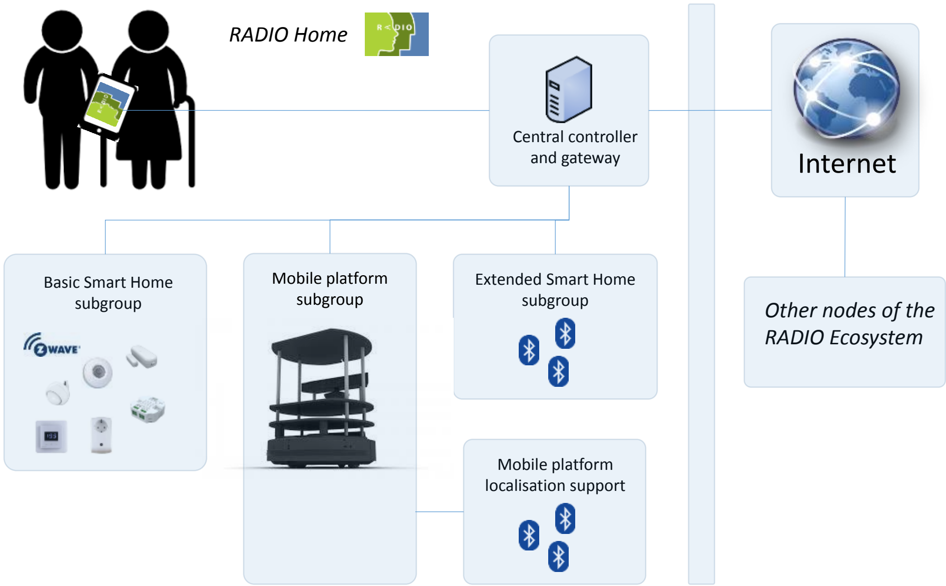

Figure 3: Conceptual architecture of RADIO Home communication channels

5.1 ADL Recognition Services

Based on our analysis of the ADL recognition methods above, we will now proceed to bundle the relevant components into services and to specify these services' information outputs and requirements. These services will be used as the building blocks of the RADIO ADL recognition architecture.

In our design we foresee the following physical infrastructure:

- acoustic sensors that integrate a microphone with Raspberry Pi
- a mobile TurtleBot2 robot that integrates microphone, depth and colour camera and on-board computer
- a Main Controller, a computer that acts as the gateway to the home and the orchestrator of the overall monitoring and reporting.

There is a single acoustic features interface which publishes a stream of triplets of feature vectors. Each message in the stream contains the current short-term frame feature vector and two mid-term rolling averages of different numbers of frames, to accommodate analyses that require deeper or more shallow acoustic contexts.

This interface was chosen because at our 50 Hz frame rate volume of traffic generated by three floating-point feature vectors is insignificant and this interface lifts the requirement to have a middleware that can latch mid-term feature vectors or synchronize mid-term and short-term feature vectors.

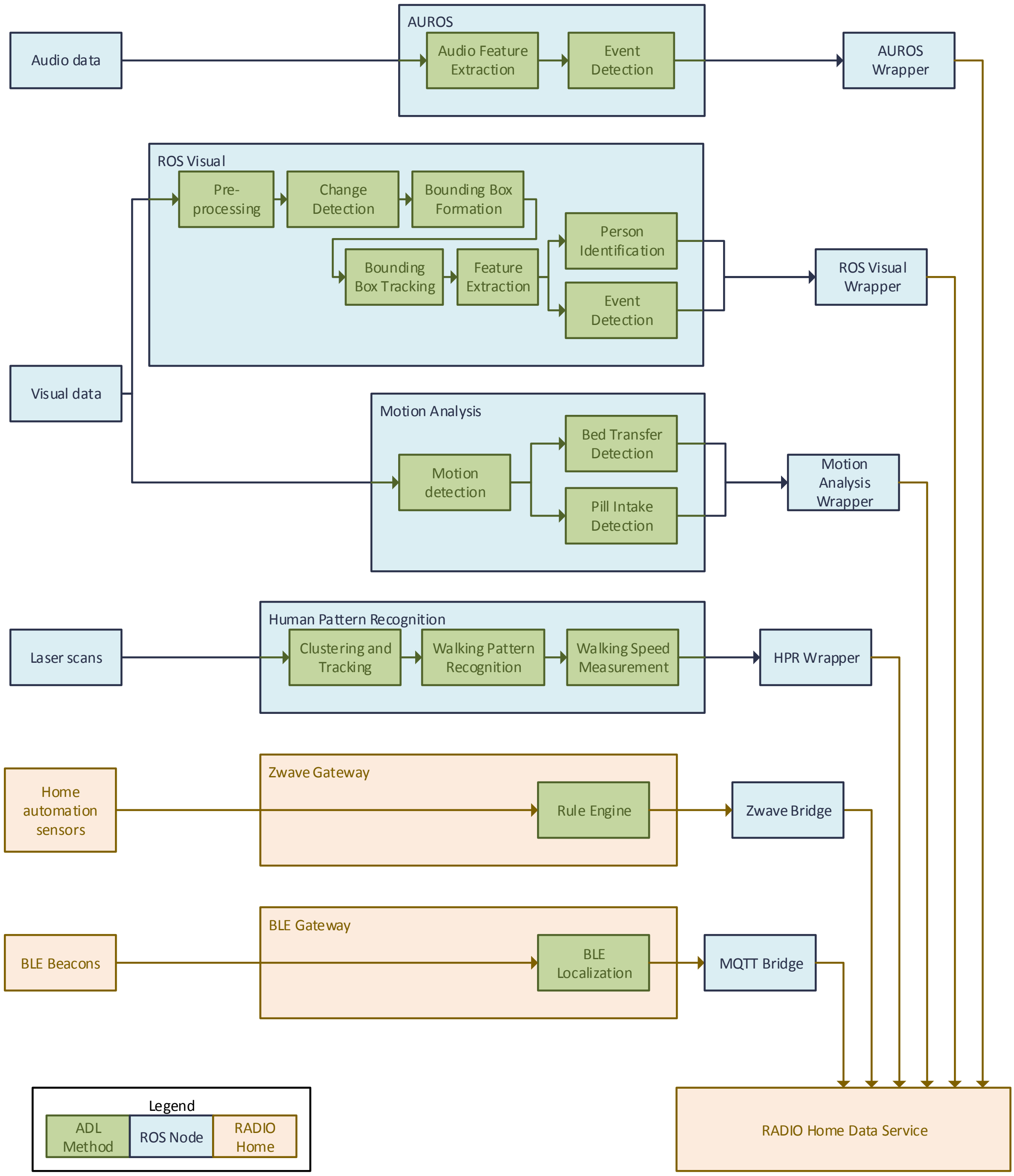

Figure 4: Conceptual architecture of ADL recognition components

Exposing the complete acquisition-feature extraction pipeline as a single service also allows us to provide a unified acoustic feature service over two heterogeneous implementations:

- The TurtleBot2 implementation comprises a microphone device driver and a feature extraction component that communicate using the ROS middleware. The service end-point is a bridge that simultaneously connects to the robot-internal ROS middleware and to the home WiFi to access robot-external services.
- The Raspberry implementation comprises a microphone device driver and a feature extraction component that communicate using MQTT. The service endpoint is a bridge that simultaneously connects to MQTT and to the home WiFi to access external services.

All instances of the acoustic features service push their vector streams to the audio pattern analysis service. This service implements unsupervised and (previously trained) supervised machine learning methods that recognize ADL events from acoustic feature vectors. The audio pattern analysis service is also distributed, with instances executing at the Raspberry and the robot's computer.

Table 6: Nature of the data exchanged between components

Component	Input	Output
Motion Analysis: Bed transfer and pill intake recognition and measurements	Image frames	Image frames, annotated with bed transfer/pill intake class.
Motion Analysis Wrapper: Logger for bed transfer and pill intake	Image frames, annotated with bed transfer/pill intake class.	Event type, annotated with timestamp and duration
ROSVisual: Recognition of movement events (bed and chair transfer, walking)	Image and depth frames	Image and depth frames, annotated with sit-to-stand class.
ROSVisual Wrapper: Measurements over movement events	Image and depth frames, annotated with sit-to-stand class.	Event type, annotated with timestamp and duration
HumanPatterRecognition: Tracks walking humans and measures walking speed	Laser scans	1sec frames, annotated with human walking annotation
HPR Wrapper: Logger of walking speed measurements	1sec frames, annotated with human walking annotation	Event type, annotated with timestamp and duration
AUROS: Audio analysis	Audio signal	500msec segments, annotated with acoustic classes
AUROS Wrapper: Audio analysis	500msec segments, annotated with acoustic classes	Event type, annotated with timestamp and duration
Rule engine	Home automation sensors	Event type, annotated with timestamp and duration
BLE Localization	BLE beacon data	Location data for BLE tagged items in environment

The vision sensing components are analogously implemented as image acquisition, feature extraction, and pattern recognition services. One divergence from the acoustic analysis case is that the graph of dependencies between vision services is not a linear progression from the content to more abstract features and events: motion detection is the only service that constantly consumes features and it triggers more complex analyses as soon as motion is detected. Furthermore, there is no single feature set that is used by all visual analyses and analyses are occasionally stacked more deeply than the features/events/ADLs layers of acoustic ADL detection.

All methods are coupled by ROS wrappers or, when on other networks/middleware than Ethernet/ROS, bridges that homogenize the output of each method to the uniform event format understood by the RADIO Home Data Service (cf. D4.2).

The conceptual architecture is also graphically depicted in Figure 4 and the interfaces between then specified in Table 6.

5.2 Concluding remarks

The RADIO Conceptual Architecture is a system of services that interact to recognize ADLs from audio-visual sensors. Our design integrates sub-systems which were originally integrated using heterogeneous middleware infrastructures. We have proposed articulation points for re-structuring these existing pipelines into a new set of services. In order to establish the right level of granularity for the functionality bundled under a single service, we used common patterns in the audio-visual analysis literature to identify services that would practically never need to be broken down into finer services.

This architecture will now be further developed in WP4, where the distribution of these components on physical computers will come into play. Another aspect that will be tacked in WP4 is the composition of the recognition services presented here into methods that use fusion to improve recognition and to recognition complex events.