AMBIENT 2011

The First International Conference on Ambient Computing, Applications, Services and Technologies

ISBN: 978-1-61208-170-0

October 23-29, 2011

Barcelona, Spain

AMBIENT 2011 Editors

Rémi Emonet, Idiap Research Institute, Switzerland

Adina Magda Florea, University "Politehnica" of Bucharest, Romania

AMBIENT 2011

Forward

The First International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT 2011), held on October 23-29, 2011 in Barcelona, Spain, was devoted for a global view on ambient computing, services, applications, technologies and their integration.

On the way for a full digital society, ambient, sentient and ubiquitous paradigms lead the torch. There is a need for behavioral changes for users to understand, accept, handle, and feel helped within the surrounding digital environments. Ambient comes as a digital storm bringing new facets of computing, services and applications. Smart phones and sentient offices, wearable devices, domotics, and ambient interfaces are only a few of such personalized aspects. The advent of social and mobile networks along with context-driven tracking and localization paved the way for ambient assisted living, intelligent homes, social games, and telemedicine.

The conference provided a forum where researchers were able to present recent research results and new research problems and directions related to them. We welcomed technical papers presenting research and practical results, position papers addressing the pros and cons of specific proposals, such as those being discussed in the standard forums or in industry consortiums, survey papers addressing the key problems and solutions on any of the above topics, short papers on work in progress, and panel proposals.

We take here the opportunity to warmly thank all the members of the AMBIENT 2011 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to the AMBIENT 2011. We truly believe that thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the AMBIENT 2011 organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success. We gratefully appreciate to the technical program committee co-chairs that contributed to identify the appropriate groups to submit contributions.

We hope the AMBIENT 2011 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in ambient computing research.

We hope Barcelona provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

AMBIENT 2011 Chairs

Advisory Chairs
Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (R. O. C.)
Naoki Fukuta, Shizuoka University, Japan
Andrzej Skowron, Warsaw University, Poland
AMBIENT 2011 Research Liaison Chairs
Rémi Emonet, Idiap Research Institute, Switzerland
Mieczysław A. Kłopotek, Polish Academy of Sciences - Warszawa, Poland

Special Area Chairs

Agents
Joseph Giampapa, Carnegie Mellon University, USA
Adina Magda Florea, University "Politehnica" of Bucharest, Romania

Context-awareness
Sylvain Giroux, Université de Sherbrooke, Canada

AMBIENT 2011

Committee

AMBIENT Advisory Chairs

Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (R. O. C.)
Naoki Fukuta, Shizuoka University, Japan
Andrzej Skowron, Warsaw University, Poland

AMBIENT 2011 Research Liaison Chairs

Rémi Emonet, Idiap Research Institute, Switzerland
Mieczysław A. Kłopotek, Polish Academy of Sciences - Warszawa, Poland

AMBIENT 2011 Special Area Chairs

Agents
Joseph Giampapa, Carnegie Mellon University, USA
Adina Magda Florea, University "Politehnica" of Bucharest, Romania

Context-awareness
Sylvain Giroux, Université de Sherbrooke, Canada

AMBIENT 2011 Technical Program Committee

Muhammad Aslam, University of Engineering and Technology - Lahore Pakistan
Flavien Balbo, University of Paris Dauphine, France
Ladjel Bellatreche, (LISI) ENSMA - Poitiers University, France
Abdenour Bouzouane, Université du Québec à Chicoutimi, Canada
Stefano Bromuri, University of Applied Sciences Western / HES-SO, Switzerland
Corneliu Burileanu, University "Politehnica" of Bucharest, Romania
Valerie Camps, Universite Paul Sabatier - Toulouse, France
Carlos Carrascosa, Universidad Politécnica de Valencia, Spain
Michelangelo Ceci, University of Bari, Italy
Keith C.C. Chan, The Hong Kong Polytechnic University -Hung Hom, Hong Kong
John-Jules Charles Meyer, Utrecht University, The Netherlands
Kuan-Ta Chen, Academia Sinica, Taiwan
Luke Chen, University of Ulster - Northern Ireland, UK
John Collomosse, University of Surrey, UK
Rémi Emonet, Idiap Research Institute, Switzerland
Klaus Fischer, German Research Center for Artificial Intelligence (DFKI) - Saarbruecken, Germany
Adina Magda Florea, University "Politehnica" of Bucharest, Romania
Naoki Fukuta, Shizuoka University, Japan
Matjaz Gams, Jožef Stefan Institute - Ljubljana, Slovenia
Joseph Giampapa, Carnegie Mellon University, USA

Sylvain Giroux, Université de Sherbrooke, Canada
Marie-Pierre Gleizes, IRIT / Université Toulouse, France
Nan-Wei Gong, MIT Media Laboratory, USA
Anandha Gopalan, Imperial College London, UK
Sandeep Gupta, Arizona State University - Tempe, USA
Vladimir Gorodetsky, St. Petersburg Institute for Informatics and Automation / Russian Academy of Sciences, Russia
Mirsad Hadzikadic, University of North Carolina - Charlotte, USA
Fumio Hattori, Ritsumeikan University - Kusatsu, Japan
Mark Hoogendoorn, VU University Amsterdam, The Netherlands
Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (R. O. C.)
Marc-Philippe Huget University of Savoie, France
Jean-Paul Jamont, University of Grenoble - Valence, France
Emil Jovanov, University of Alabama in Huntsville, USA
Ioannis A. Kakadiaris, University of Houston, USA
Anthony Karageorgos, Technological Educational Institute of Larissa - Karditsa, Greece
Pavel Kisilev, Hewlett-Packard Laboratories - Haifa, Israel
Mieczyslaw A. Klopotek, Polish Academy of Sciences - Warszawa, Poland
Franziska Klugl, Örebro Universitet, Sweden
Matthias Klusch, German Research Center for Artificial Intelligence (DFKI) - Saarbruecken, Germany
Don Kraft, Louisiana State University - Baton Rouge, USA
Satoshi Kurikara, Osaka University, Japan
Freddy Lécué, University of Manchester, UK
Ioan Alfred Letia, Technical University of Cluj-Napoca, Romania
Henrique Lopes Cardoso, DEI/FEUP, Portugal
Jun Luo, Shenzhen Institutes of Advanced Technology/Chinese Academy of Science, China
Ana Martinez Enriquez, CINVESTAV, Mexico
Zulfiqar Ali Memon, Sukkur Institute of Business Administration - Sind, Pakistan
Paolo Merialdo, Università degli studi Roma Tre, Italy
Marie-Francine Moens, Katholieke Universiteit Leuven, Belgium
Gearóid O'Laighin, National University of Ireland - Galway, Ireland
Helen Paik, University of NSW, Australia
Alina Pommeranz, Technische Universiteit Delft, The Netherlands
Yacine Sam, Université François-Rabelais Tours, France
Majid Sarrafzadeh, UCLA, USA
Patrick Sayd, CEA, France
Shishir Shah, University of Houston, USA
Andrzej Skowron, Warsaw University, Poland
Yehia Taher, Tilburg University, The Netherlands
Lei Tang, Yahoo! Labs, USA
Surapa Thiemjarus, Sirindhorn International Institute of Technology / Thammasat University, Thailand
George Vouros, University of the Aegean - Samos, Greece
Hoi-Jun Yoo, KAIST, Korea
Markus Zanker, Alpen-Adria-Universität Klagenfurt, Austria
Yuan-Ting Zhang, The Chinese University of Hong Kong (CUHK), Hong Kong
Ingo Zinnikus, German Research Center for Artificial Intelligence - Saarbrücken, Germany

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Context-aware Multimodal Feedback in a Smart Environment

Didier Perroud, Leonardo Angelini, Elena Mugellini, Omar Abou Khaled
Department of Information and Communication Technology
University of Applied Sciences of Western Switzerland
Fribourg, Switzerland
{didier.perroud, leonardo.angelini, elena.mugellini, omar.aboukhaled}@hefr.ch

Abstract - **The use of multimodality improves interaction between the user and the computer. Particularly, the use of multimodal feedback within a smart environment facilitates the integration of technology into the daily activities of the user. However the choice of the suitable output modalities requires knowledge of the user context to be effective. This paper presents an approach for the generation of context-aware multimodal feedback in the context of ambient intelligence. Our solution is based on the NAIF Framework, which handles the creation and management of a smart environment. A preliminary prototype has been developed and tested in order to validate the proposed approach.**

Keywords - multimodal feedback; multimodal fusion; context-aware; smart environment; ambient intelligence; ubiquitous computing; NAIF Framework

I. INTRODUCTION

The technological development of last years has considerably changed our daily life. Many electronic devices populate our environment and their use has become a habit. It is practically impossible to imagine a day without using a mobile phone, a computer or a television. The use of electronic devices covers most of our activities, whether related to work, entertainment or learning. The technical maturity of production means for electronic components allows devices to be more powerful, smaller in size and equipped with an increased number of functionalities. Latest devices have several on-board sensors that allow improvements of the usability. Thanks to them the local context is taken into account during software development. For example, a smart phone can now detect its orientation to adapt the graphical representation of an application [1], can take into account light condition to change the luminosity of the screen and use head proximity sensors to turn off the display when answering a call.

Miniaturization and reduced costs of electronic components encourages manufacturers to integrate multiple communications interfaces in devices. This integration extends the functional capabilities of each device. For example, nowadays it is possible to find on the market televisions that allow direct access to Internet. The widespread diffusion of communication means and the ubiquitous sensors integration open the door to the exchange of information between devices in a given environment. The increased connectivity and information exchange between systems are the basis for developing novel, intelligent applications. The primary purpose of these intelligent applications is to provide a greater control of the environment to the people. This field is called *Ambient Intelligence*, as described by P. Remagnino et al. [2], with some typical scenarios presented in [3].

The daily presence of people in an ambient intelligent environment involves a change of habit in terms of interaction. The smart environment must be non-intrusive and able to understand user's needs. The distribution of systems in the space also requires the use of another way of interaction than standard mouse and keyboard. Indeed, the user must be able to interact with its environment without constraints on the devices to use. Multimodality seems to be a suitable and flexible solution for the interaction between the user and a smart environment.

The work presented in this paper is related to multimodal generation of output content, taking into account user context in an intelligent environment composed of autonomous distributed systems. Issues concerning the creation of an intelligent environment will not be discussed in this paper. The approach proposed for the multimodal output generation uses NAIF (Natural Ambient Intelligence Framework) [13], which allows the setup of an intelligent environment.

The paper is organized as following. Section II presents some work in the scientific community addressing the concepts of multimodal fission and multimodal generation with context management. NAIF Framework is briefly presented in Section III. The proposed approach to context-aware multimodal feedback generation is explained in Section IV. Section V presents a prototype that validates the use of the proposed approach. Section VI concludes the paper and discusses future work.

II. RELATED WORK

The use of multimodality in human-computer interaction is a subject frequently discussed in the scientific community. The acquisition of multiple signals from the human is the first big challenge to solve in order to build systems more comfortable for the user. Nevertheless, the output generation should be also investigated to grant multimodality in both directions of the human-computer interaction. Few projects have dealt with multimodality in both input and output processes, the most related to our work are presented in the following paragraph. To the best of our knowledge no one treated input and output

 ISBN: 978-1-61208-170-0

multimodality using context information from a smart environment.

SmartKom [4] is a multimodal communication system that combines voice, gesture and facial expression as input and output. The aim of this project is to create a natural experiment of communication between user and machine. The concept is based on a virtual agent capable of interpreting communicative intentions in the context of assistance to purchase a ticket. This project is relatively complex since it deals with multimodality in a full and symmetrical spectrum. The generation of multimodal output of SmartKom is based on a system developed under another project called WIP [5][6]. The main component of WIP is a presentation planner that transforms the intentions of communication in presentation tasks. Then the planner allocates these tasks to specific generators of modality like voice or gesture. The fission engine of SmartKom uses therefore an important knowledge database that contains all different patterns and presentation strategies available for each modality. SmartKom is an example of project that addresses the challenges of multimodality as input and output. Its development is centered on the user and multimodality is limited to a few specific modalities. The system SmartKom cannot be applied in a distributed environment like a smart environment.

C. Rousseau et al. [7] proposed a conceptual model called WWHT for the multimodal presentation of information. The model WWHT is articulated around four basic concepts, *What*, *Which*, *How* and *Then,* describing the life cycle of a multimodal presentation adapted to the context of ongoing interaction. The presentation process is based on 4 components: the information to present, the interaction components of the system, the ongoing context of the interaction and the resulting multimodal presentation. A first semantic fission of information occurs in the step *What* in order to form smaller units of information. In the next step, *Which,* the various units of information are allocated to a specific modality (e.g., visual or haptic) and the choice of communication medium is made (screen, sound speaker). During this step, the dependencies between allocated modalities are also assessed. Multimodality is addressed in this assessment by considering the CASE model [8], which classifies the possible combinations of modalities. The generation of modalities occurs in step *How* where the interaction components of the system produce outputs. Finally the step *Then* is responsible of context monitoring in order to adapt multimodal presentation.

POPEL [9][10] is a generating component of natural language integrated in the XTRA Framework. XTRA offers a treatment of multimodality as input and output. Supported communication channels are focused on natural language and complementarity by gestures. Like WWHT, Popel separates the information fission process of output generation. The first step *POPEL-WHAT* aims at the selection of information to be transmitted depending on the current context. *POPEL-HOW* is responsible for generating output. WWHT and POPEL use a concept of division between the modality and its representation. This separation makes the model of WWHT or the implementation of POPEL much more flexible. It also helps break down the complexity of multimodality processing.

W3C Multimodal Interaction Framework [11] is a specification provided by W3C to extend the Web for supporting several modes of interaction. The specification addresses the multimodal interaction as input and output. It describes the different components that any multimodal system must implement. The internal operations of components are not included in the description. The W3C specification separates the generation from the rendering of multimodality. The exchange of information between components is also specified. It consists of several markup languages using the eXtensible Markup Language (XML) specification of W3C. For example, the component audio rendering can handle a Speech Synthesis Markup Language (SSML) document while a graphics rendering component can interpret an eXtensible HyperText Markup Language (XHTML). The use of a standardized language to exchange information improves the modularity of the framework. This concept is particularly suitable in a distributed environment context.

The DynAMITE project exploited multimodality in a smart environment where heterogeneous devices can interoperate thanks to a common framework. This work deals with the ubiquitous computing within dynamic ad-hoc devices ensembles. Even if multimodality is addressed at both input and output sides, the context information of the environment is not exploited in this project. Unfortunately the project DynAMITE seems to have been abandoned. Some explanations about the project in general or about the internal components can be found in [14][15][16].

Our approach presented in this paper focuses on the context-aware multimodal generation within an intelligent environment. The execution context of the applications is a distributed environment, where multiple applications can run in parallel. Therefore, our approach must be as flexible as possible because each application requires different needs in terms of generation of the multimodal feedback. Our approach uses some concepts of the aforementioned works that improve flexibility and modularity. We separate the notion of modality and representation. We also use markup language to exchange information within the environment. Moreover our framework aims to address the issue of a context-aware multimodal feedback on multiple distributed systems in a smart environment.

III. NAIF FRAMEWORK

The generation of multimodal output presented in this paper takes place in a smart environment. Our approach is based on NAIF [12], which is an acronym for *Natural Ambient Intelligent Framework*; it is developed since 2009 at the University of Applied Sciences of Western Switzerland, in Fribourg. This framework aims to address the issues related to the setup and development of a smart environment. This section is a brief presentation of the framework; further details are available in [13].

 ISBN: 978-1-61208-170-0

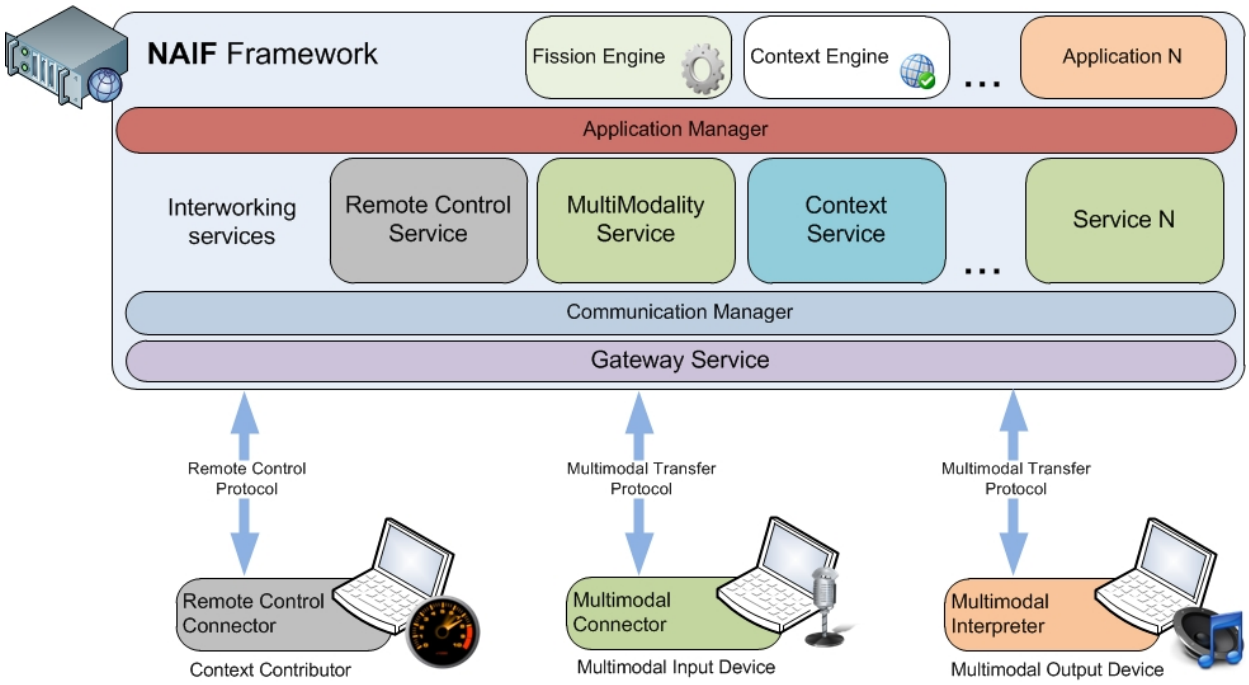

Figure 1: Context-aware multimodal generation architecture based on NAIF

The basic idea of NAIF is that our everyday environment is full of devices performing specific tasks independently. Each of these devices has inherent capabilities that can be shared with other systems of the environment. Therefore, the devices can benefit from the presence of other systems to do their own tasks or accomplish a common goal. For example, a smart phone could use a radio in another room to broadcast the ringing of a phone call when the user is not present. In the same way, the smart phone could collaborate with an application of energy saving by sharing onboard light sensor measures to signal light on.

NAIF is built on client-server architecture. A central platform provides services to devices that constitutes smart environment. The operation of NAIF is based on three concepts described below.

1) Intercommunication

Each system of the smart environment hosts a *software agent*. This agent manages the communication with the central platform and the data exchange format. On the server side, a service called *Gateway* is responsible for routing communication between systems. To support heterogeneity, the communication protocol consists of XML frames. The protocol NAIF is therefore located at application level and relies on a TCP stack.

2) Interworking services

To ensure interoperability between systems in the environment, the central platform offers on-demand services. These services are mandated to make collaboration between systems effective but also to facilitate the work of developers that want to exploit the advantages offered by the available systems. For example, NAIF provides a service called *Remote Control Service* that shares measures of sensors available on the devices of the environment. The service layer of the platform is extensible. Developers can add services as needed.

3) Multimodal interaction

NAIF proposes an approach to multimodal interaction in smart environments. An interworking service called *MultiModality Service* is provided by the central platform. This service operates as a directory that allows sharing modalities of systems equipped with appropriate devices and software, which is charged to process signals from onboard sensors. For example, a system with a speech recognition engine and a microphone can share the detection of words pronounced by user. This service deals with modalities only as input and has no mechanism of fusion.

The following section presents our context-aware multimodal feedback concept integrated in NAIF.

IV. CONTEXT-AWARE MULTIMODAL GENERATION

As discussed above, the presence of a user in a smart environment involves new ways of interaction. This problem is often approached from the point of view of combining and exploiting different input interaction modalities. The concept presented in this paper focuses on combining and exploiting different *output* interaction modalities. As the user benefits from multimodality as a natural channel to communicate with the environment, the environment must be able to produce a multimodal feedback to the user. The generation of feedback should occur under the best conditions possible to improve comfort for the user. To find the best conditions, the user context should be considered. Context data will allow the smart environment to produce an optimal multimodal feedback. For example, if the user has a visual impairment, an auditory feedback should be produced instead of a visual one.

 ISBN: 978-1-61208-170-0

The concept of context-aware multimodal generation explained in this section is an extension of NAIF. The extensibility of the framework allows the inclusion of new interworking services. The solution detailed in this paper is therefore based on the integration of new services and software components. Figure 1 shows the architecture of NAIF with extensions. The following sections describe this architecture extension.

A. Multimodal Input Sharing

As previously explained, NAIF includes a *MultiModality Service*, which allows the sharing of input modalities. The idea behind this concept is based on the collaboration between systems in a smart environment. Imagine a TV, it requires a remote control to understand user commands. If a system in the environment has a voice recognition engine, the TV could benefit from it to improve its interaction input channels. The current version of the *MultiModality Service* operates as a directory. Systems capable of processing input signals from users publish their functionalities in the directory. Systems wishing to receive a modality may register to the directory. When a modality is detected, e.g., a word is spoken, a notification will be sent to all systems that observe it.

The concept of modality is very subjective. For example, it is difficult to characterize what a voice modality is: it can contain just a word, a phrase, an intonation or a pronunciation. NAIF avoids this problem by allowing developers of different applications to model their own modality according to their needs. Each modality will be described in an XML schema. Schemas will be shared with all systems of the smart environment using the *MultiModality Service* through a software component called *Multimodal Connector*. The *Multimodal Connector* component is hosted on the environment systems that use the *MultiModality Service*. The *Multimodal Connector* is responsible of managing the link between the system and the interworking service. It uses the *Multimodal Transfer Protocol* to communicate with the service. This XML protocol is encapsulated within the data field of NAIF frames. The structure of this protocol is simple. It contains only the transfer of commands like *publish* a modality, *register* to a modality or *notify* a modality. As the schema of each modality is available in the connector, systems can interpret the contents of the frames received.

B. Multimodal Output Generation

The generation of multimodal feedback in our framework is designed in order to allow a system that has limited feedback capabilities to use the outputs of other systems. A heating controller for example will be able to use a TV screen to display an alert when resources are missing. The designed approach extends the possibilities of the *MultiModality Service*. By adding a new *publish* command, a system can offer its feedback capabilities to other systems. The service can now receive two types of commands. The first announces a system as a supplier of modality, the second as a provider of feedback. This extension requires a change in the *Multimodal Connector* and a protocol arrangement. A generate-modality message to feedback generator systems is added in the *Multimodal Transfer Protocol*. This change raises a new problem. Generating systems have to interpret this modality message to produce feedback. To solve this problem, a new software component called *Multimodal Interpreter* is placed on the published multimodal feedback generator systems.

The multimodal interpreter is responsible for the transformation of a modality in an effective presentation of the information. This component depends on the local platform. It cannot be totally generic. The concept of modality in the *MultiModality Service* is not limited for input; the developer can specify its own modality for output as well, if needed. The interpreter must however be able to understand modalities that are sent for generation. In summary, if a system sends text as voice modality, the work of the *Multimodal Interpreter* is to synthesize the voice through text-to-speech.

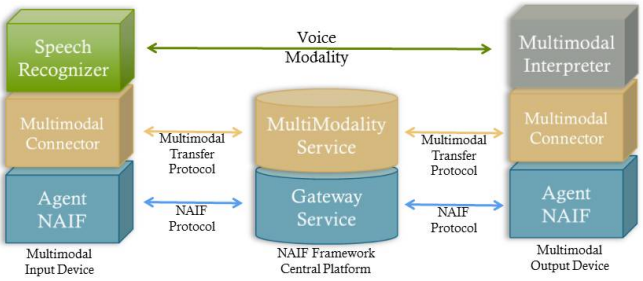

Figure 2: Software stack of an exchange of voice modality between a speech recognizer system as input and a published multimodal output system.

The proposed extension involves the creation of roles as shown in the architecture in Figure 1. A system can act as a multimodal input device. In this case it only hosts the *Multimodal Connector* and shares its modalities. A system can also act as multimodal output device. In addition to the connector, the output system hosts the *Multimodal Interpreter* as shown in Figure 2 and generates received modalities. Input and output roles can be combined.

C. Context management

In our work, we have defined the context as the description of the parameters surrounding an action. As part of a smart environment, the context concerns users as well as all the environmental parameters that affect a given action. In terms of multimodal feedback generation, the context is a primordial factor that can disrupt interaction with the user. A context management should be present to produce a good quality of feedbacks.

Our approach is based on the integration of a contextual reasoning engine and an interworking service in NAIF. The *Context Engine* is located at the application layer of the framework. This layer allows applications to run on the central platform and to access interworking services through an *Application Manager*. The role of *Context Engine* is to extract the context state from available data of the smart environment. The management of the context adds a new role for the systems of the environment. Systems that act as *Context Contributor* publish data from their sensors, or from their local context. A smart phone for example can publish its current direction sensor value or its current activity. The publication of data sensors is available in NAIF through the *Remote Control Service*. The devices can use the *Remote Control Connector* that links to the service on the central platform. Therefore the

 ISBN: 978-1-61208-170-0

context engine can use sensor data available in the *Remote Control Service* to extract a global context. The engine then builds a representation of the context that is stored in the new service called *Context Service*. This service operates as a repository of the current state of the smart environment. Each system of the smart environment can access this service to obtain information about the ambient context. This allows preserving the autonomy of the different systems within the environment.

A first version of the *Context Service* has been developed. However, at the moment, only the software components running on the central platform can access the *Context Service* (the other systems in the environment cannot access this service because no connector component exists). An effective representation of a context must also be studied. This modeling process is also a subjective task, which could limit the functionality of the framework. One might ask if the context and the reasoning engine should not be a single entity. To maintain the flexibility of NAIF, the *Context Service* has been separated from the context-reasoning engine. This allows to easily change the engine with another that implements a different algorithm.

D. *Multimodal fission*

The generation of multimodal feedback is effective through the *MultiModality Service*, thus the framework should be able to split information into multiple modalities. Fission can be approached from two aspects. The first approach is the fission of information semantics in order to extract multiple information units, which will be processed into modalities. This task of division is extremely complex and requires an important research work. Instead the concept of fission addressed in our approach concerns the choice of the best modalities for a specific feedback that has to be sent to the user.

We introduced a component called *Fission Engine* in the application layer of the central platform. The engine receives feedback intentions as input, and then takes care of electing the best modalities and the best systems of the environment to generate these feedbacks. The Fission Engine relies on context state to select modalities and systems. It consults the *Context Service* directly to obtain the current context state. Insofar as the choice requires additional reasoning, we can imagine that the *Context Engine* offers facilities to solve the problem of election. The *Fission Engine* communicates directly with the *MultiModality Service*. Upon receiving a feedback intent, it chooses the best modalities and then it sends a feedback generation message to the selected systems. Take the example of a security application that controls the entrance of a room. When an alert message should be sent, the application formulates its intent to the *Fission Engine*, which will then select the best modalities on the best feedback generator systems. Taking into account the context of the user's proximity, for example, the *Fission Engine* could ask the generation of visual and audio alert on the systems closest to the user.

Like the reasoning engine of the context, there are many possible approaches to implement the *Fission Engine*. The intelligence of the engine can fundamentally modify the effectiveness of feedback. The management of multimodality is also reflected in the engine. The CASE properties [8] can be used in the election procedure to improve feedback. The *Fission Engine* is positioned in the application layer on the central platform. Therefore this software component can be exchanged easily and the flexibility of NAIF is preserved.

V. PROTOTYPE

A prototype has been developed in order to demonstrate the feasibility of the concept. The objective of the prototype is to show that it is possible to exploit the capabilities in terms of modality generation of different autonomous systems in a smart environment, taking into account the state of the ambient context. The prototype must test the different software components presented in the previous sections. The architecture of this prototype is shown in Figure 3.

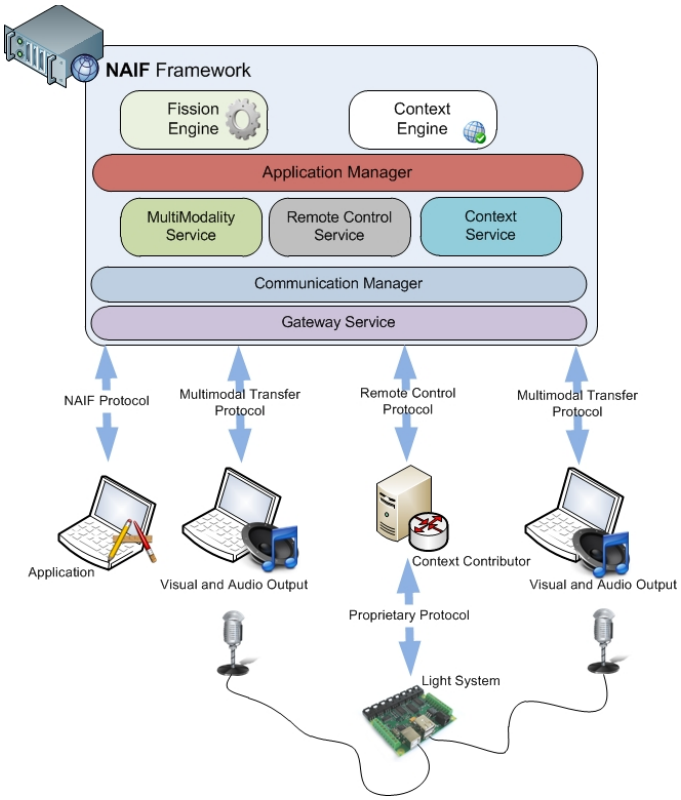

Figure 3: Prototype architecture

Two laptop placed in the environment are able to produce auditory (text-to-speech) and visual feedback. Both systems publish their generating capacity of two modalities in the *MultiModality Service*. Two sensors of noise intensity are located near these multimodal output devices. A computer processes the signals from these sensors and publishes values in the *Remote Control Service*. A *Context Engine* accesses data from *Remote Control Service* and built a simple XML representation (e.g. *<Noise location ='1'>57</Noise>*) of the sound context that it stores in the *Context Service*. By subscribing to notifications of the *Remote Control Service,* the *Context Engine* performs regular updates of the context representation.

An application running on a third laptop with no audio output requires the display of a text and a vocal commentary as

 ISBN: 978-1-61208-170-0

feedback. Using its NAIF communication agent, it sends feedback intent to the *Fission Engine*. Its frame is structured as follows:

<Feedback><Text>textual feedback</Text>
<Voice>vocal feedback</Voice></Feedback>

The *Fission Engine* queries the *MultiModality Service* to find available output generators and it analyzes the XML representation in the *Context Service.* Then, it chooses the modalities and selects the best feedback generators according to the noise state of the context. Finally the *Fission Engine* sends modality generation requests to the selected systems through the *MultiModality Service*.

Feedback generation requests are received in the *Multimodal Interpreter* of the selected output devices. Modalities are ultimately generated by the respective hardware. In this prototype, only the sound context is taken into account. The intelligence of *Fission Engine* is very simple too. Nevertheless, the voice generation and display of text are produced on the system the least disturbed by noise.

The results obtained are consistent with the objectives. The integration of our concept in NAIF let us take advantage of the flexibility of the Framework. The application could for example make the fission itself and directly contact *MultiModality Service* to send a generation request using the *Multimodal Connector*, or it could use the *Fission Engine* as explained in this prototype.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a context-aware approach to multimodal feedback generation in a smart environment. The context analysis is included in the solution to improve the efficiency of multimodal generation. Our proposal is designed as an extension of the NAIF Framework, which allows setting up, and developing an intelligent environment.

The contribution of our approach could help the development of intelligent environments. By using multimodal feedback, ambient intelligence can communicate with the user trough more natural channels. Therefore, the integration of smart environments in everyday human life could be less complicated. Based on the context, ambient intelligence can also react to disturbances that might limit the usability of the global system. Our contribution relative to the NAIF Framework provides new features for developers of applications in smart environments. It is now possible to benefit from the intrinsic characteristics of autonomous systems in terms of output. The developer can also focus on the user when designing new applications because of the full spectrum of the multimodal communication with the user. Our extension therefore complies with the conception of the Framework.

The prototype presented in this paper validates the proposed context-aware fission concept, however some improvements are still needed. The reasoning engine of the context should be improved with algorithms and technologies that will be used to build the representation of the context. An approach based on ontologies, for example, could meet the needs of our concept.

Finally, the use of the *Fission Engine* must be improved, mainly in the formulation of feedback intentions. This concept should be modeled to establish the communication protocol used to contact the *Fission Engine*. The concepts of modalities and feedbacks should be further investigated

VII. REFERENCES

[1] S. Valbert and C. Per Thorsø, " Improving usability of mobile devices by means of accelerometers ", Master Thesis, Kongens Lyngby, 2009, ISBN: IMM-M.Sc.-2009-32.

[2] P. Remagnino and G.L. Foresti, " Ambient intelligence: a new multidisciplinary paradigm ", IEEE Xplore, vol. 35, 2005, pp. 1-6, doi: 10.1109/TSMCA.2004.838456 .

[3] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J-C. Burgelman, " Scenarios for ambient intelligence in 2010 ", ISTAG report, 2011, p. 54.

[4] W. Wahlster, " SmartKom : fusion and fission of speech, gestures and facial expressions", Proc. 1st International Workshop on Man-Machine Symbiotic Systems, Kyoto, Japan, 2002, pp. 213-225.

[5] W. Wahlster, E. André, W. Finkler, H.-J. Profitlich, and T. Rist, " Plan-based integration of natural language and graphics generation ", Artificial Intelligence 63, 1993, pp. 387-427, doi: 10.1016/0004-3702(93)90022-4

[6] E. André, W. Finkler, W. Graf, T. Rist, A. Schauder, and W. Wahlster, " WIP: the automatic synthesis of multimodal presentations ", Intelligent Multimedia Interfaces, AAAI Press, Menlo Park, 1993, pp. 75-93, ISBN: 0-262-63150-4

[7] C. Rousseau, Y. Bellik, and F. Vernier, "WWHT: un modèle conceptuel pour la présentation multimodale d'information", Proc. IHM, 2005, pp. 59-66, doi: 10.1145/1148550.1148558.

[8] L. Nigay and J. Coutaz, " Design space for multimodal systems: concurrent processing and data fusion", Proc. Conference on Human Factors in Computing Systems (INTERACT '93 and CHI '93), ACM, 1993, New York, pp. 172-178, doi: 10.1145/169059.169143

[9] D. Schmauks and N. Reithinger, " Generating multimodal output: conditions, advantages and problems ", Proc. 12th conference on Computational linguistics (COLING 88), Budapest, Hungry, 1988, pp. 584-588, doi: 10.3115/991719.991758

[10] N. Reithinger, " POPEL—a parallel and incremental natural language generation system ", In C. Paris, W. Swartout and W. Mann, ed, Natural Language Generation in Artificial Intelligence and Computational Linguistics, Kluwer, Dordrecht, Netherlands, 1991.

[11] http://www.w3.org/TR/mmi-framework/ 01.08.2011

[12] http://www.naif-project.ch 01.08.2011

[13] D. Perroud, F. Barras, S. Pierroz, E. Mugellini, and O. Abou Khaled, " Framework for development of a smart environment, conception and use of the naif framework ", Proc. 11th International Conference on New Technologies of Distributed Systems (NOTERE 11), Paris, France, 2011, pp 151-157.

[14] http://www-past.igd.fraunhofer.de/igd-a1/projects/dynamite/ 01.08.2011

[15] K. Richter and M. Hellenschmidt, " Interacting with the ambiance: multimodal interaction and ambient intelligence ", Architecture, vol. 19, 2004, p. 20, 2004.

[16] M. Hellenschmidt and T. Kirste, " SODAPOP: a software infrastructure supporting self-organization in intelligent environments ", Proc. Industrial Informatics (INDIN04), Berlin, 2004, pp. 479 – 486, doi: 10.1109/INDIN.2004.1417391

 ISBN: 978-1-61208-170-0

Context-Aware 3D Gesture Interaction Based on Multiple Kinects

Maurizio Caon[1,2], Yong Yue[1]
[1]Faculty of Creative Arts, Technologies & Science
University of Bedfordshire
Luton, United Kingdom
e-mail: {maurizio.caon, yong.yue}@beds.ac.uk

Julien Tscherrig[2], Elena Mugellini[2], Omar Abou Khaled[2]
[2]Department of Informatics
University of Applied Sciences of Western Switzerland, Fribourg
e-mail: {julien.tscherrig, elena.mugellini, omar.aboukhaled}@hefr.ch

***Abstract*—This paper presents a novel context-aware system for deictic gestures interaction with smart environments. The system tracks multiple users; moreover, it recognizes inhabitants' postures and gestures in real-time. This information, enriched with smart objects coordinates, is reconstructed in a 3D model to allow the recognition process. Finally the system executes the programmed tasks to support the users' activity. Two Microsoft Kinect depth cameras have been used to acquire the data and a framework for the communication with the smart objects has been adopted. A first prototype has been developed and an evaluation test with 13 users has been conducted in order to assess the usability of the system. Results show that this interaction experience has been really appreciated by the users.**

Keywords-Ambient Intelligence; Smart Environment; Gesture Interaction; Posture Recognition; Depth Cameras

I. INTRODUCTION

Norman's invisible computer [1] and Weiser's ubiquitous computer [2] theories led to the conception of Ambient Intelligence (AmI). This multidisciplinary paradigm is intended to create new smart infrastructures that seamlessly integrate intelligent services [3]. This new conception of computing is thought to be invisible but always active in the background to grant all the services that are supposed to be necessary to the user. This novel anthropomorphic human-machine model of interaction permits the user to move into the foreground in complete control of the smart, augmented environment which interprets actions to support and enhance the abilities of its occupants in executing tasks [4]. Such a system has awareness about the user's current activity, situation and intention before the activity is actually completed to provide appropriate support. A technology that allows reading the human mind does not exist yet, and capturing actor's intention implicitly is a very hard challenge [5]. For this reason, a smart environment cannot limit its decisions to the elaboration of context information, but it has to allow the inhabitants to interact with the environment in order to explicit their intentions and goals. According to the current trends, the design of an interactive environment interface aims especially to solve important issues related to the usability and the adaptiveness of such interfaces. In order to create an end-user friendly interface, many researchers are focused on natural ways of interaction as speech and gestures [6]. In particular, deictic gestures play an important role since they are intuitive and commonly used by humans to reference objects and devices by pointing at them [7]. Therefore, deictic gestures are really significant for human-environment interaction. On the other hand, correct deictic gestures interpretation by the system depends on the user context (e.g., position and orientation in the 3D space) and this involves the need of a situational awareness about the smart objects placed in this interactive space.

A real-time context-based system for deictic gestures interaction with smart environments is presented in this paper. This system grants human actors to interact with a smart environment pointing at the smart objects. The context information comes from the data related to the states and positions of the smart objects, and to the tracked inhabitants' postures. All this information is modeled in a 3D virtual space. The sensors that are used to acquire the data are two Microsoft Kinect depth cameras. Hence the user does not need to wear any special device.

The 3D camera technology is positioned to become ubiquitous; in fact the Microsoft Kinect is a really cheap off-the-shelf device which can provide quite accurate depth information (11 bit data for 2,048 levels of sensitivity) at a good frame-rate (30 Hz). The research community has already manifested strong interest in this new device, also for applications that go far beyond simple video-gaming [8][9][10].

The rest of the paper is organized as following. In Section 2, the related work is presented and the scenario is described in Section 3. The tests and the considerations related to the use of multiple Kinects are discussed in Section 4; the architecture of the system is described in Section 5. Section 6 presents the tests that have been made in order to evaluate the system. Section 7 is dedicated to the conclusion reporting also the future work.

II. RELATED WORK

Interaction between human beings and smart environments is a real demanding research area [6]. Gestures are really significant for this kind of applications [11].

Many research works are focused on deictic gestures and often they prefer cameras as sensors to capture the inhabitants' movement information, as in [12] and [13].

 ISBN: 978-1-61208-170-0

Information extraction from 2D video streams involves many limitations because pointing in a real room needs 3 dimensions for a complete representation. In fact, the authors of [14] adopted stereo-cameras to extract depth information from disparity map.

Postures recognition is another research domain that captured researchers' interest [15]. Most of the works adopted video-cameras as sensing device [16][17][18]. The elaboration of data deriving from a single 2D video flow involves many problems, e.g., occlusion and cluttered background, and puts many limits in recognizing human postures. Indeed, Chu and Cohen adopted a system based on four synchronous cameras to recognize postures and gestures [19]. Another solution to add important spatial information to recognize postures consists in using a depth camera, as in [20].

Depth cameras can really improve system performances for the interaction in a smart environment, as Wilson and Benko demonstrated in [21]. On the other hand, this kind of systems allows the interaction only with selected surfaces.

In this paper, a novel system that recognizes 3D deictic gestures and human postures using multiple depth cameras is presented. It reconstructs context information elaborating spatial coordinates of the smart objects (that are previously inserted in the system) and of the inhabitants (that are constantly tracked in the 3D space), combining them with the deictic gestures and postures data to improve the interaction experience.

III. SCENARIO

Youngblood et al. defined a smart environment as one that is able to acquire and apply knowledge about the environment and its inhabitants in order to improve their experience in that environment [22]. Designing a smart room that can achieve this goal involves the context awareness and the possibility of interaction with the people. Gestures are a natural way of interaction for humans and integrating these commands with information coming from the situation can make the environment to support user's tasks. A smart room that can achieve this goal has to recognize the inhabitants' activity, it has to understand the direct commands ordained by the users and it must integrate many smart objects to communicate with. Therefore, the smart environment detects the human posture and the smart objects state; indeed it can create context information. Commands, acquired from the users present in the environment, are interpreted referring to the previously modeled context. Our context information comes from the data related to the states and positions of the smart objects, and to the tracked inhabitants' postures.

The target scenario deals with a smart living room which recognizes deictic gestures and human postures, tracks the inhabitants and allows them to interact with smart objects that are present in the interactive space, see Figure 1. The novel approach of this system assigns different meanings to the pointing gesture according to the user's posture. If the user is sitting on the couch in front of the TV and points at the media center, then the TV is turned on. If the user is standing in the center of the room and he points at the media center, then the radio is turned on (the environment is set to interpret this situation as the user would like to listen to music). The spatial coordinates and the postures are crucial for the interpretation of the meaning of the gestures performed by the user, but the state of the smart objects is important as well. If the human points at the lamp, which is turned off, then the system turns it on; on the contrary if the lamp is already on, the system turns it off. Another application deriving from the posture recognition part is aimed to emergency purposes. When the system detects a person lying on the ground for a period superior to a guard time that has been previously set, the system calls the rescues. On the other hand if the human is lying on the sofa, the smart environment closes the blinds and turns off the lights to permit to the user to rest comfortably.

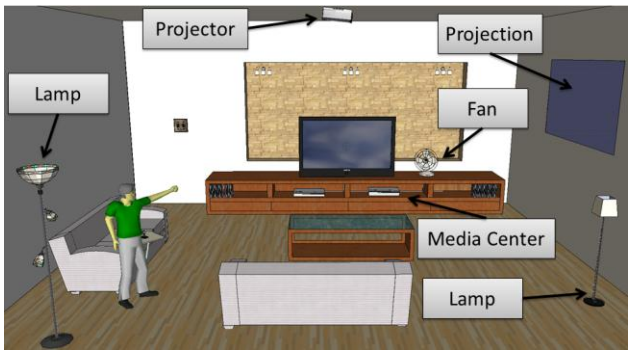

Figure 1. Scenario.

IV. USING MULTIPLE KINECTS

Robust interactive human body tracking has many applications, in particular it can be important in human-computer interaction; depth cameras can simplify reaching this task and the Microsoft Kinect represents the first down-market device that can allow capturing 3D general body motions and shapes at interactive rates [20]. However, using multiple Kinects involves interference between the infrared laser patterns that are at the base of the functioning of this device. Each Kinect projects its own infrared pattern for the calculation of the depth information and interferences can degrade the information quality creating black spots on the 3D image. In order to assess if the interferences change significantly referring to the number of active Kinects and their positions, 5 different configurations have been tested. Figure 2 represents the camera configurations that have been tested. The Kinects have been positioned in A, B and C. The colored triangles represent the field of view of the cameras from the A, B and C positions. The striped areas of the triangles represent the interactive areas, or rather, the areas where people can be easily tracked. This area begins at a distance of 0.8 m from the Kinect and arrives to 3.5 m. A person was present in the test scenario and he was positioned on the white circle in the center of the figure. The A, B, and C positions are at the same distance from the person, in order that the points of the patterns projected from the infrared lasers have same brightness and dimensions on the person. The optical axis of the Kinect positioned in A intersects the optical axis of the Kinect

 ISBN: 978-1-61208-170-0

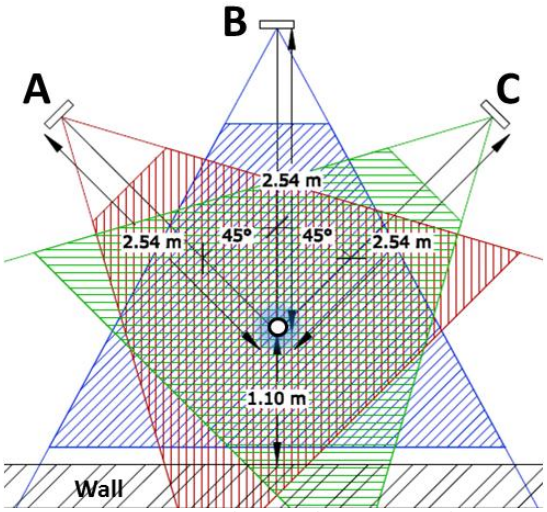

Figure 2. Representation of the interference test configurations.

positioned in B forming an angle of 45°. The optical axis of the Kinect positioned in A intersects perpendicularly the optical axis of the Kinect positioned in C. In configuration 1 there was only one active Kinect and it was positioned in A. In configuration 2 there were two active Kinects and they were both positioned in A. In configuration 3 there were two active Kinects, one was positioned in A and the other one in B. In configuration 4 there were two active Kinects, one was positioned in A and the other one in C. In configuration 5 there were three active Kinects, one was in A, one in B and one in C. In Figure 3, the captures for every configuration taken from the Kinect that has always been in A are reported. The black pixels in the captures represent the pixels without depth information.

To quantify the interference effect, the number of pixel without depth information has been calculated. Since the pixels without depth information change during time also on a static scene, then this number has been calculated making an average on 1000 frames for every configuration. The depth sensor of the Kinect captures 640x480 pixel frames; therefore, every frame has got 307200 pixels with depth information. For the configuration 1, an average of 5325 pixels without depth information has been calculated (with standard deviation of 165 pixels); for the configuration 2 the average was of 14018 pixels and the standard deviation was of 404 pixels; for the configuration 3 the average was of 12502 pixels and the standard deviation was 319 pixels; for the configuration 4 the average was of 13000 pixels and the standard deviation was of 295 pixels; for configuration 5 the average was of 21813 pixels and the standard deviation was of 432 pixels. After these tests, we verified that the interference caused by two Kinects is not significant for the skeleton tracking and it remains almost constant regardless the relative position of the two cameras. However, using two Kinects in configuration 4 permits capturing the tracked users' movements from very different perspectives. This configuration permits to capture a very big portion of the users' bodies avoiding in many cases the occlusion of some

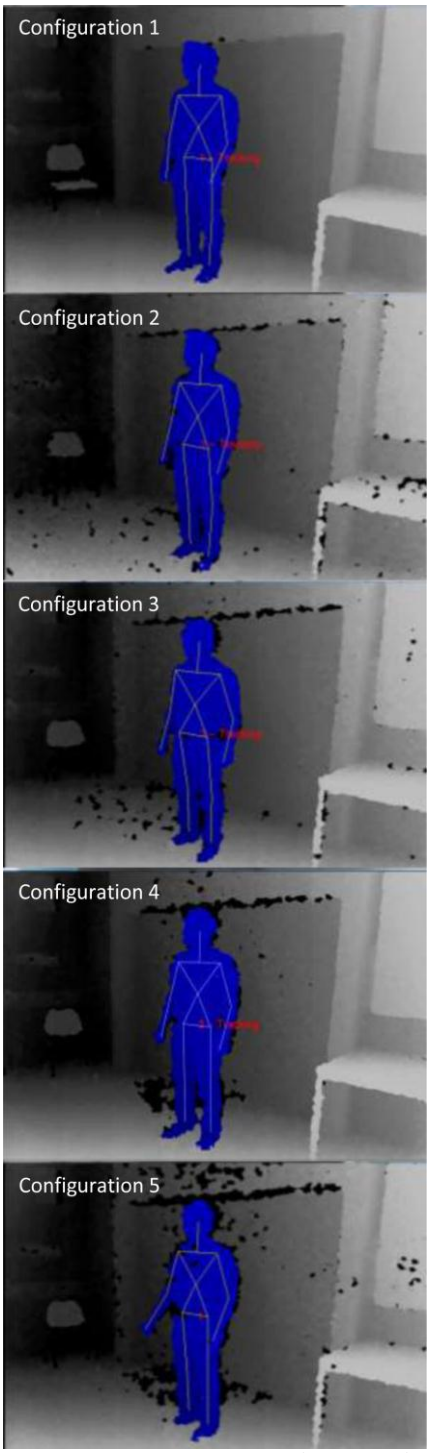

Figure 3. Scene captures during the interference test for every configuration.

limbs. In Figure 4 a), an example is shown; on the upper part the Kinect can capture the left side of the user and cannot see the legs; in the lower part of this figure there is the view from the other Kinect present in the system that can capture the information about user's legs, but it cannot track his left arm. The system combines the data coming

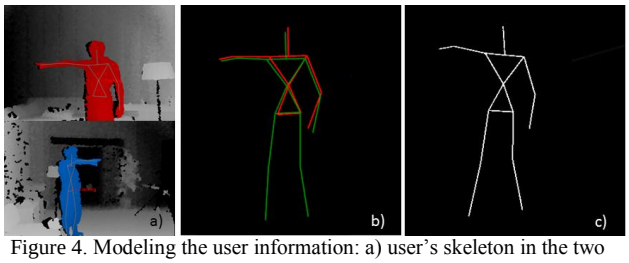

Figure 4. Modeling the user information: a) user's skeleton in the two Kinect views; b) 3D model of the user's skeletons captured by the two Kinects; c) 3D fusion of the user's two skeletons in one skeleton

from the two Kinects to reconstruct the whole user's skeleton as explained in the next section.

Using three Kinects in configuration 5 doubles the number of the pixels without depth information generated by the interference but it does not add significant information for the user's skeleton reconstruction. More tests with three and four Kinects surrounding the users will be executed soon.

According to the results of these tests, in our system we decided to use two Kinect cameras positioned as in configuration 4. The developed system is presented in the next section.

V. SYSTEM ARCHITECTURE

The developed system has been designed to be modular. It consists of two acquisition modules (one for each Kinect camera) and a central module.

The acquisition modules are based on the OpenNI libraries [23] and can track the people present in the vision area. Each module constructs a skeleton model of each tracked user that will be represented in a specific XML structure. These data will be sent to a central module for the 3D modeling. These modules communicate with XML messages using the UDP protocol. Every XML message contains the information about the coordinates of every joint of every tracked person (that is uniquely identified) and the ID number of the Kinect camera that sent these data. Moreover, in the XML message there is also the destination IP address since this system has been designed to communicate with other machines in order to make possible a future distributed version to spread the global interactive area.

The central module makes the fusion of the data concerning the same tracked user to create a 3D skeleton, as shown in Figure 4. This 3D skeleton is calculated with coordinates of the different joints and it is placed in the 3D model of the environment. The fusion algorithm calculates the difference between the coordinates of every joint, and then it makes an average of these coordinates weighting the information of the more reliable data. In fact, the system assigns a higher weight to the coordinates that come from the Kinect capturing the highest number of joints information. Depending on the user's position and posture, a Kinect can see a bigger or smaller part of the user's body. When a Kinect cannot track a specific part of the body to calculate the joints coordinates, it does not send any information to the central module about that joint. For this reason, the algorithm for the fusion assigns a higher weight to the Kinect that can detect more joints. When the coordinates of a specific joint are provided from only one Kinect because the other one cannot track this body part, the algorithm uses directly the unique received data. When both Kinects cannot provide the coordinates of a specific joint, then the system ignores that joint waiting for new data.

The data related to the coordinates of the smart objects present in the interactive area must be inserted in the 3D model of the environment using a dedicated interface of the central module. The advantage of this data representation consists in the possibility of setting spatial constrains in 3 dimensions for the interaction with the smart objects.

A. *Calibration*

The cameras calibration is crucial to reconstruct a 3D model using simultaneously multiple depth cameras. The coordinates of the joints coming from the two cameras must be represented in the 3D reconstruction of the environment. In order to find the right relative coordinates of the points captured from the two different points of view, a transformation matrix must be determined. The transformation matrix in 3D is a special 4x4 matrix and is based on quaternions [24]. This transformation matrix provides a rotation around the x, y and z axis. The calibration phase aims to calculate the 16 values of this matrix. To solve this matrix it is necessary to obtain the coordinates of 4 fixed points from the two cameras. The resolution of quaternion matrix aims to resolve the transformation point from a relative coordinated system to the main coordinated system.

Considering four common points on the two different Kinect cameras with known exact coordinates, then it is theoretically possible to calculate the transformation matrix.

The calculation of the transformation matrix based on quaternion aims to resolve 16 equations with 16 unknowns. All the calibration process is effectuated from the central module.

The cameras calibration phase must be effectuated during the set-up of the system. The calibration permits to position the two Kinects in every desired configuration, the only constrain is that they must capture the same scene. In fact, the two Kinects must see at least four common points to accomplish the calibration. This system has been programmed to register up to 10 common points to calculate several transformation matrixes. Afterwards, the system computes for each transformation matrix the average delta between the main coordinates and the transformed coordinates of all the captured points; therefore, the transformation matrix with minimum average delta is chosen. The system can utilize the calculated transformation matrix as long as the Kinects remain in the same positions.

B. *Gesture and Postures Recognition*

The 3D model of the environment includes the users' skeletons and the smart objects. The system recognizes the users' postures and pointing gestures from the coordinates of the joints in real-time. This recognition process is based on simple conditions referred to some values of the joints and

 ISBN: 978-1-61208-170-0

the relative distances between them. The postures that are recognized by the system are two: standing and sitting. This information is completed with the relative positions of the objects in the 3D model of the environment. When the joints of the arm assume specific values, the system projects a prolongation of the arm and calculates if it intersects the active area attributed to a specific object. Our system has been integrated in the NAIF framework [25]. NAIF framework handles the creation and management of a smart environment. It manages the set-up and the communication between smart objects and devices present in the environment. Thanks to NAIF, our system can check the current object state and generate the suitable command, e.g., if the lamp state is off then it sends the appropriate command to turn it on.

VI. System Evaluation

In order to have a feedback about the system usability and to understand the limitations of our prototype, we performed an evaluation test composed of two phases. The subjects of this test are 13 users (9 men and 4 women) with different backgrounds and origins, and with age between 19 and 28 years.

A. First Phase

The subjects have been conducted to the smart living room where a simple scenario has been prepared. One user at a time has been asked to enter in the room and to interact with the system (see Figure 5). After the skeleton tracking initialization stage (the user has to remain in a pose for few seconds in front of each Kinect device), the user had to point at a lamp to turn it on; afterwards the user had to point at the media center to turn on the radio and later he had to do it again to turn it off. Afterwards, he had to sit down on the couch and to point at the media center to turn on the TV. The system never failed the gesture or the posture recognition during the test. Once finished the interaction session in the smart living room, every subject evaluated the experience through a System Usability Scale (SUS) [26] questionnaire rating the system features according to a 5-point Likert scale. The statements covered a variety of aspects of system usability, such as the need for support, training, complexity, efficiency (how much effort is necessary in achieving those objectives) and experience satisfaction.

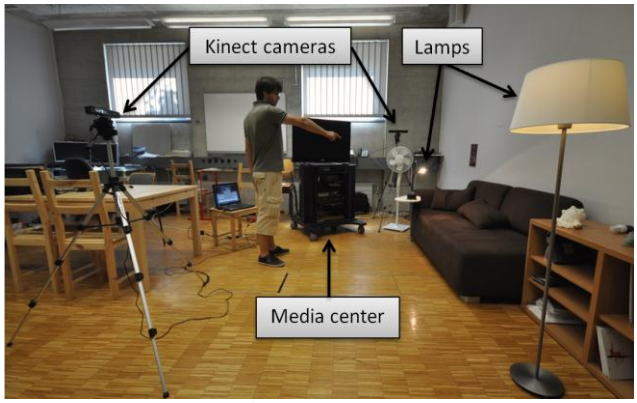

Figure 5. One of the users testing the system.

The users' evaluations assessed the system usability as excellent with an average SUS score of 90.6 points and a standard variation of 5 points.

B. Second Phase

This phase consisted in an interview where the users have been asked to express their impressions and suggestions. Most of them said that the skeleton tracking initialization stage could be really annoying for an everyday interaction in a real smart room. The subjects have been asked to say if they have missed the voice interaction modality in this test scenario and everybody answered negatively, moreover they expressed their appreciation about this interaction modality through deictic gestures. Some of the users remarked that they would like also other gestures to go beyond the turning on or off the household appliances, e.g., they would like to interact with the media center to change TV program or the volume.

Another limitation that came from our analysis is the pre-determined set of tasks that the system executes referring to the users' gestures and postures. In fact, a system that automatically learns user's habits could be preferable to a programmed one. Therefore, in order to make this system more human-centered, the integration of learning algorithms has been thought in order that the system can learn users' habits.

VII. Conclusion and Future Work

In this paper, a real-time context-based system for deictic gestures interaction with smart environments has been presented. The system tracks multiple users and reconstructs situational information collecting data about the people's postures and coordinates. Moreover, this software realizes a 3D model of the environment where only the tracked users and the smart objects are present. The users' skeletons are modeled referring to the joints coordinates captured by two calibrated Microsoft Kinect cameras; the smart objects coordinates have been inserted previously in the system and their current states are provided by NAIF framework. The 3D model of this information makes the postures and deictic gestures recognition easy. The context awareness makes possible to interpret the pointing gesture referring to the posture and coordinates of the user, giving different meanings to the same gesture in order to execute different tasks. In the current prototype the tracked people can point at two lamps and at the media center. Pointing at the lamps turns them on or off (it depends from the previous state). Pointing at the media center turns on or off the radio if the user is standing, otherwise turns on or off the TV if he is sitting on the couch. The usability tests assessed that this interaction modality with the smart environment is really intuitive; indeed the users do not need training to interact with the smart objects and they affirmed that they had a really pleasant experience. Future works are already planned in order to shorten, or if possible, eliminate the skeleton tracking initialization stage (resulted annoying for the users

 ISBN: 978-1-61208-170-0

during the evaluation tests) and to add the learning algorithms for a more human-centered system that learns users' habits. Afterwards, more gestures for an augmented environment control will be implemented. Adding more gestures will increase the recognition complexity and the precision of the Microsoft Kinect could become critical, for this reason a comparison with a ground truth will be conducted. Finally, the system will be tested with multiple users interacting with the environment at the same time.

REFERENCES

[1] D. A. Norman, The Invisible Computer, Cambridge, MA: MIT Press, 1999.

[2] M. Weiser, The Computer for the 21st Century, Scientific American, September 1991.

[3] N. Shadbolt, "Ambient intelligence," IEEE Trans. Intell. Transp. Syst., vol. 18, no. 4, pp. 2–3, Jul.–Aug. 2003.

[4] P. Remagnino and G.L. Foresti, "Ambient Intelligence: A New Multidisciplinary Paradigm," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 35, Jan. 2005, pp. 1-6.

[5] D. Surie, T. Pederson, F. Lagriffoul, and L. E. Janlert, "Activity Recognition using an ' Egocentric ' Perspective of Everyday Objects," Time, 2007, pp. 1-16.

[6] D. Cook and S. Das, "How smart are our environments? An updated look at the state of the art," Pervasive and Mobile Computing, vol. 3, Mar. 2007, pp. 53-73.

[7] M. Karam, "A taxonomy of gestures in human computer interactions," 2005, pp. 1-45.

[8] E. Suma, D. Krum, B. Lange, S. Rizzo, and M. Bolas, "Faast: The flexible action and articulated skeleton toolkit," IEEE Virtual Reality, 2011, pp. 247-248.

[9] C. Schönauer, T. Pintaric, and H. Kaufmann, "Full body interaction for serious games in motor rehabilitation," Proc. 2nd Augmented Human International Conference, ACM, 2011, p. 4.

[10] A. DeVincenzi, L. Yao, H. Ishii, and R. Raskar, "Kinected conference: augmenting video imaging with calibrated depth and audio," Proc. ACM 2011 conference on Computer supported cooperative work, ACM, 2011, p. 621–624.

[11] A.M. Rahman, M.A. Hossain, J. Parra, and A. El Saddik, "Motion-path based gesture interaction with smart home services," Proc. of the seventeen ACM international conference on Multimedia (MM '09), 2009, p. 761.

[12] R. Kehl and L. Van Gool, "Real-time pointing gesture recognition for an immersive environment," Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 577-582.

[13] A. D. Wilson and a F. Bobick, "Recognition and interpretation of parametric gesture," Proc. Sixth IEEE International Conference on Computer Vision, 1998, pp. 329-336.

[14] E. Seemann and R. Stiefelhagen, "3D-tracking of head and hands for pointing gesture recognition in a human-robot interaction scenario," Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 565-570.

[15] E. Farella, A. Pieracci, L. Benini, and A. Acquaviva, "A Wireless Body Area Sensor Network for Posture Detection," Proc. 11th IEEE Symposium on Computers and Communications (ISCC'06), 2006, pp. 454-459.

[16] S. Mu and I. Lii, "A multiscale morphological method for human posture recognition," Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 56-61.

[17] N. Zouba, B. Boulay, F. Bremond, and M. Thonnat, "Monitoring activities of daily living (adls) of elderly based on 3d key human postures," Cognitive Vision, 2008, p. 37–50.

[18] L.B. Ozer and W. Wolf, "Real-time posture and activity recognition," Proc. Workshop on Motion and Video Computing, 2002, pp. 133-138.

[19] C.W. Chu and I. Cohen, "Posture and gesture recognition using 3D body shapes decomposition," Human-Computer Interaction, 2005.

[20] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," In CVPR, 2011.

[21] A.D. Wilson and H. Benko, "Combining multiple depth cameras and projectors for interactions on, above and between surfaces," New York, New York, USA: ACM Press, 2010.

[22] G.M. Youngblood, D.J. Cook, L.B. Holder, E.O. Heierman, "Automation intelligence for the smart environment," Proc. International Joint Conference on Artificial Intelligence, 2005.

[23] http://www.openni.org/ 07.08.2011

[24] B.K.P. Horn, "Closed-form Solution of Absolute Orientation using Unit Quaternions," Journal of the Optical Society of America, Series A, 4, 4, 1987, pp. 629-642.

[25] D. Perroud, F. Barras, S. Pierroz, E. Mugellini, and O. Abou Khaled, "Framework for development of a smart environment, Conception and Use of the NAIF Framework," Proc. 11th International Conference on New Technologies of Distributed Systems (NOTERE), Paris, France, 2011.

[26] J. Brooke, "SUS-A quick and dirty usability scale," Usability evaluation in industry, 1996, pp. 189–194.

 ISBN: 978-1-61208-170-0

A Model for Activity Recognition and Emergency Detection in Smart Environments

Irina Mocanu and Adina Magda Florea
Computer Science Department
University "Politehnica" of Bucharest
Bucharest, Romania
e-mail: irina.mocanu@cs.pub.ro, adina.florea@cs.pub.ro

***Abstract*—Activity recognition has become an important feature in smart environments designed for helping old people living alone and independently in their homes. This paper presents a model for detecting emergencies in case of human activity recognition in such a smart environment. The emergencies detection is performed using a stochastic context-free grammar with attributes together with a domain activity ontology for modeling the daily programme of the supervised person.**

Keywords-activity recognition; smart environments; context-free grammar with attributes; emergency detection.

I. INTRODUCTION

The percentage of elderly people in today's societies keeps on growing. As a consequence we are faced with the problem of supporting older adults in loss of cognitive autonomy who wish to continue living independently in their home as opposed to being forced to live in a hospital. Smart environments have been developed in order to provide support to the elderly people or people with risk factors who wish to continue living independently in their homes, as opposed to live in an institutional care.

In order to be a smart environment, the house should be able to detect what the occupant is doing in terms of one's daily activities. It should also be able to detect possible emergency situations. Furthermore, once such a system is completed and fully operational, it should be able to detect anomalies or deviations in the occupant's routine, which could indicate a decline in his abilities.

Our goal is to develop an integrated ambient intelligent system for elderly people or people with risk factors, called AmIHomCare, which includes several modalities of surveillance and assistance of people with special needs. One component of the system is the human activity recognition module, which is dedicated to recognizing human activities as part of daily living. Identification of daily activities is done using ontologies [1], vision-based models [2], sensor-based models [3] or different algorithms such as Hidden Markov Models [4]. Also the module is responsible with the detection of emergency situations during the daily activities. The paper presents a model for emergency detection for a set of human activities known to take place in the living room. The rules used for activities recognition are modeled with a stochastic context-free grammar. Different types of emergencies are detected during the activity parsing using a set of data: (i) an attribute for modeling the duration of activity added in the activity grammar and (ii) a domain ontology for represented activities' properties (e.g., normal duration, usual place of performing an activity, time of the day or usual sequence of some activities).

The rest of the paper is organized as follows: Section II presents existing methods for activity recognition and Section III describes the general structure of the AmIHomCare system in which the component responsible for activity recognition will be integrated. Section IV introduces the model for emergencies detection in activity recognition. Experimental results are presented in Section V and the paper is concluded in Section VI.

II. RELATED WORK

Activity recognition became an important research issue related to the successful realization of intelligent pervasive environments. It is the process by which an actor's behavior and his or her environment are monitored and analyzed to infer the activities. Activity recognition consists of: (a) activity modeling, (b) behavior and environment monitoring, (c) data processing and (d) pattern recognition. There are several approaches for activity recognition as described in [1]:

- Vision-based activity recognition which uses visual sensing facilities: camera-based surveillance systems to monitor an actor's behavior and the changes in its environment. It is composed of four steps: human detection, behavior tracking, activity recognition and high-level activity evaluation. Various other research approaches used different methods such as: single camera or stereo and infra red to capture activity context. For example, in [2], a system for recognizing activities in three steps is presented: (i) track pedestrians across a scene and recognize a set of human activities; (ii) use multiple cameras to observe the scene and (iii) use an active dual camera for task recognition at multiple resolutions.
- Sensor-based activity recognition uses sensor network technologies to monitor an actor's behavior along with its environment. In this case there are sensors attached to humans. Data from the sensors are collected and analyzed using data mining or machine learning algorithms to build activity models and perform activity recognition. In this case, there recognized activities included human physical

 ISBN: 978-1-61208-170-0

movements: walking, running, sitting down/up as in [3]. Most of wearable sensors are not very suitable for real applications due to their size or battery life.

- Activity recognition algorithms can be divided in three categories: machine learning techniques, grammar based techniques and ontological reasoning. Many types of machine algorithms for activity recognition were developed, including Hidden Markov Models, Bayesian Networks or Support Vector Machine techniques. Among them Hidden Markov Models and Bayesian Networks are the most commonly used methods in activity recognition. Standard Hidden Markov Models (HMM) are employed for simple activity recognition as described in [4][5]. They are not suitable for modeling complex activities that have large state and observation spaces. Parallel HMM [6] are proposed for recognizing group activities by factorizing the state space into several temporal processes. Another approach for activity recognition based on HMM is semi-Markov models as described in [7]. Other models are represented by conditional random fields. Conditional random fields are discriminative models for labeling sequences [8]. They condition on the entire observation sequence, which avoids the need for the assumption of independence between observations. The Conditional Random fields method performs as well as or better than the HMM even when the model features do not violate the independence assumptions of the HMM as described in [8]. Other type of methods for human activity recognition is based on hierarchical Bayesian networks [9] or dynamic Bayesian networks [10], which can model the duration of an activity. Another method for activity recognition uses context-free grammar. A context-free grammar for the description of continued and recursive human activities is presented in [11]. Other types of grammars are used for activity recognition. An example is stochastic context-free grammars. Detection and recognition of temporally extended activities and interactions between multiple agents are modeled using a probabilistic syntactic approach (stochastic context- free grammar), as described in [12]. Stochastic context-free grammars are also used for recognizing kitchen-specific activities as in [13]. In [14], an attribute grammar which is capable of describing features that are not easily represented by finite symbols is proposed. This grammar is used for representing multiple concurrent events which involve multiple entities by associating unique object identification labels with multiple event threads. The ontological reasoning models activities in an ontology and the reasoning is realized based on the properties of the compound entities as described in [1].

III. ACTIVITY RECOGNITION IN AMIHOMCARE SYSTEM

The general structure of an ambient intelligent system (an intelligent house) for home medical assistance of elderly or disabled people, called AmIHomCare is presented in [15]. The paper describes the main components of the system, the purpose of each component and the links between them.

The main objective of this system is to develop an intelligent environment for ambient assisted living, which achieves home monitoring and assistance for elderly people or patients with risk factors, controls the environment, and detects medical emergencies.

The system has four main components:

- A component to monitor and control ambient factors such as light, temperature, humidity, as well as home security;
- A component to monitor patient health status by using non-intrusive and intrusive sensors, and send alerts in case of risk values;
- A component to achieve patient gesture recognition and gesture-based interaction with a “robot like” personal assistant;
- A component to achieve human activity monitoring (the supervising system), offers to the patient pervasive access and retrieval to medical products information (the retrieval system). Both the supervising system and the retrieval system work based on captured images and patient specific context.

AmIHomCare also includes a connection to a call center and a home assistance center. The AmIHomCare system proactively assists people in their daily activities or medical needs, detects medical emergencies, and sends information to a call center.

The supervising system analyses the images captured by the supervision cameras. For each image, the context of the detected person together with its pose are determined. The context together with the pose form a sub-activity. An activity is composed by a set of successive sub-activities. The process for human activity recognition is described in Figure 1 as it is presented in [16].

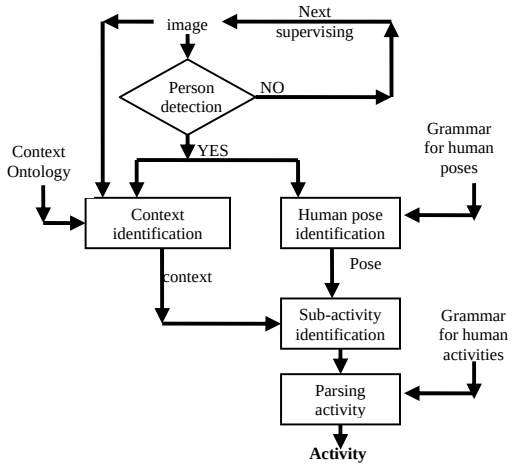

Figure 1. Human activity recognition

A supervising camera is installed in each room of the house. It takes snapshots at a predefined interval. Each image is analyzed in order to perform human activity recognition. The main steps of the human activity identification, as described in [16] are: (i) Person detection in the image; (ii) Identification of the detected person's context (iii) Person's pose identification in the recognized context and (iv) Sub-activity identification based on the obtained context and on the person's pose. The context of a person refers to the zone from the room in which the person was detected, together with the surrounding objects (furniture objects) from the room. Thus, each room is divided in zones of interest. For example, the living room from Figure 2 is divided in three zones: R1 (the resting zone), R2 (the reading zone) and R3 (the dining zone). For example, if we consider a person in the living room from Figure 2 in zone R2, near the armchair, his context will be (zone:R2, furniture object: armchair).

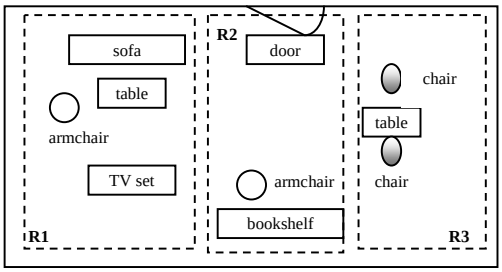

Figure 2. Areas from the living room.

The image from the supervising camera is analyzed in two steps in order to detect the person's context. First the image is annotated so that image objects will get associated keywords. Image annotation is performed with a parallel genetic algorithm, described in [17], which determines the best match between each image region and the corresponding objects in the room. Secondly, the image is analyzed for person detection. The bounding boxes of the objects from the image (the detected person and the furniture objects) are compared. Thus the objects close to the supervised person will be identified. Then these objects will be used to determine the zone of the room in which the person is located. This assumes a model of the house is available. The house model is a domain ontology – the context ontology, which consists of the rooms in the house, the zones inside each room and the furniture objects. Each room has its component zones and each zone consists of the component furniture objects. For a better person's context identification, each furniture object from the house and the supervising person have associated a depth using a distance sensor (sonar). Thus each furniture object from the ontology will have an associated list of one or more depth values. Each depth value in the list will be measured considering a known angle of the distance sensor. Next the ontology is queried with the objects close to the supervised person. The query result is zone R which contains the majority of objects close to the detected person (under a predefined threshold). The zone R together with the nearest furniture object(s) from R forms the person's context.

The position of the detected person is obtained using its depth (from the distance sensor) combined with a movement sensor. In case of movement detection, the human pose is identified. The human pose identification is described in [16]. The human body is modeled by its body components. A list of known poses stored in an and-or graph is used for this purpose. The pose is obtained using a set of rules modeled as a stochastic context-free grammar, which is transformed into the equivalent and-or graph. The human pose identification is performed probabilistically, by bottom-up constructing a list of parse trees in the grammar. The parse tree with the biggest probability is chosen from the result list.

Each activity is decomposed into a sequence of sub-activities. Each sub-activity consists of the human pose and its context. For the human activities, in [16] there are considered three possible activities:

- Watch TV: walk through the living room and sit down on the armchair (from R1) or on the sofa;
- Have a snack: walk through the living room and sit down on a chair near the dining table;
- Read a book: walk through the room, go to the bookshelf and sit down on the armchair (from R2).

A set of successive sub-activities are assembled in an activity using a stochastic context-free grammar.

IV. EMERGENCIES DETECTION IN HUMAN ACTIVITY RECOGNITION

Activities of Daily Living (ADL) have some general characteristics: (i) they are composed of component sub-activities (for example for preparing a meal somebody takes the ingredients and then will cook it); (ii) they are performed in specific circumstances (in a specific environment with specific objects for specific purposes) (for example people go to bed in a bedroom at a specific time or meals are made in a kitchen with a cooker) and (iii) people have different lifestyles. Thus activity recognition must be used together with an emergency detection technique. For this purpose the activity recognition technique must be used together with the daily programme of the supervised person which will be inferred with an emergency detection mechanism.

Emergencies detection in activity recognition implies (i) activity recognition based on a model for activities and second (ii) emergency detection inferring the detected activity together with the general daily programme of the supervised person. Thus it is necessary to have a model for activities such that it can be easy to detect some abnormalities in the daily programme of the supervised person.

In this paper, four types of daily activities abnormalities are considered:

- Activities with a longer duration than usual, which will generate a duration emergency;
- Activities which are performed in a wrong place, which will generate a context emergency;
- Activities which are performed at an unusual time of the day, which will generate an unusual time emergency.

 ISBN: 978-1-61208-170-0

- Activities which are performed in an unnatural order, which will generate an unnatural order emergency.

In all the above cases, we can have a high or a low emergency.

The majority of the methods for human activity recognition are based on a set of rules for discovering activities. These rules are modeled as Bayesian networks or stochastic context-free grammars, as described in Section II. In these cases the time of the day, the place or the order of the activities are ignored. In [10], the duration of an activity is also taken into account, leading to an activity model based on dynamic Bayesian networks. In this case, the modeling of the four abnormalities described above: (duration, place, time of the day and order of activities) is done with a stochastic context-free grammar with an attribute for activity recognition, plus an activity ontology. The attribute from the grammar will keep the duration for the recognized sub-activity/activity. The normal duration, the time of the day, place and activity order are modeled by the activity ontology (the daily programme ontology).

The use cases described in [18] are used for abnormalities modeling. Mary is an old woman who is living alone in the smart environment. The following examples are considered:

a) Reading a book: Mary is going to the bookshelf. She is taking a book, then she is walking to the armchair. But after that she will be walking through the room for a long time. This situation is a low emergency and the smart environment will alert Mary that she wants to read a book and she will go to the armchair which is near the bookshelf and sit down.

b) Resting on the sofa: Mary is resting on a sofa for a period which is longer than usual. In this situation for the beginning it is a low emergency and the smart environment will alert Mary to stand up. If Mary doesn't stand up, the emergency will become a high emergency and the smart environment will send a message to the call center to intervene.

c) Lying down on the floor: Mary is lying down on the floor in the middle of R2 zone. In this situation it is a low emergency and the smart environment will alert Mary that she must wake up. If she will remain lying down on the floor the emergency will became high and the smart environment will send a message to the call center.

d) Having a nap after the breakfast in the morning: Mary just finished her breakfast and she went to the bedroom and fell asleep. This is unusual, because after she finished her breakfast she made a phone call with her friend. In this situation, initially a low emergency is triggered and the smart environment will alert Mary that she must wake up. If she will remain falling asleep the emergency will became high and the smart environment will send a message to the call center.

For doing this it is necessary to modify the model for activity recognition so that the duration of a sub-activity/an activity can be determined. Thus the stochastic context-free grammar used for activity recognition in [16] will be modified in a stochastic context-free grammar with an attribute. The attribute associated with the grammar will model the duration of each sub-activity / activity. The normal duration of a sub-activity, the place in which an activity is performed, time of the day at which each activity is realized and the acceptable sequence of activities form the daily programme of the supervised person. The daily programme of each supervised person is modeled in a domain activity ontology. A domain activity ontology for some activities of daily living was created. The daily programme of each supervised person will be an instance from this ontology. The conceptual activity model of the ontology is described in Figure 3.

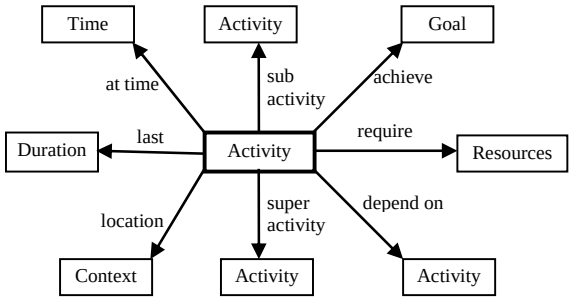

Figure 3. The conceptual activity model from the daily programme ontology

Each activity is described by a number of properties: (a) *Sub-activity*: some activities are composed of a set of component sub-activities (for example the reading activity is composed of: walking – going to the bookshelf, taking the book and then walking again – to sit on the armchair); (b) *Super-activity*: each activity may be part of other activity (for example walking is part of the reading activity or of the resting activity); (c) *Context*: each activity is performed at a specific location (context) (for example the resting activity is performed on the sofa in the living room or on the bed in the bedroom); (d) *Duration*: each activity has a normal duration; (e) *Time*: each activity is performed at a specific time of the day; (f) *Goal*: each activity has a specific goal; (g) *Resources*: some activities need a set of resources (for example the reading activity requires a book taken from the bookshelf); (h) *Dependence*: each activity is made after another activity or after a set of activities (for example resting takes place after lunch).

The activity recognition and emergency detection can be described as in Figure 4. As described in Section III, first the context of the supervised person has been identified using the context domain ontology; then the human pose is estimated using a stochastic context-free grammar (SCFG) combined with an and-or graph which describes a set of known human poses. The identified context and the human pose form a sub-activity. Successive sub-activities will be assembled in a known activity by parsing in a stochastic context-free-grammar with an attribute (ASCFG). The ASCFG is obtained from the stochastic context-free

grammar used in [16] for activity recognition in which is added an attribute for the duration for each sub-activity.

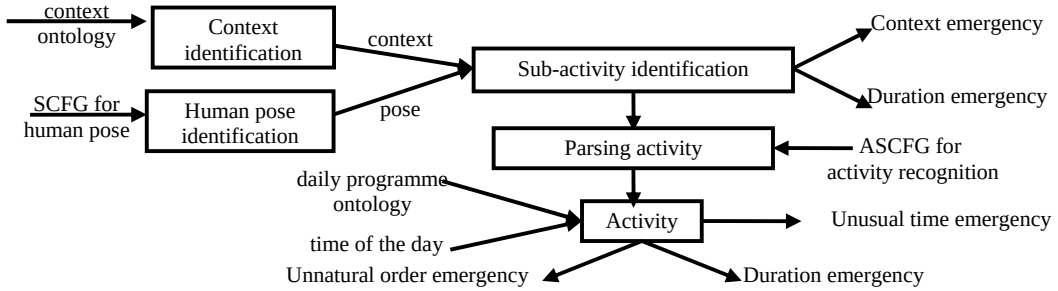

Figure 4. The general structure of the proposed model for the activity recognition- emergency detection

A part of the productions of the stochastic context-free grammar with attributes for the reading activity from the R2 zone is described in Figure 5:

- The rule (1) of the grammar consists of the three analyzed activities: reading, having a snack and watching TV. The duration of the activity is the duration of one of the possible activities.
- The rest of the rules (2) – (7) represents the rules for the reading activity.

(1) Start → Reading$^{0.333}$ | HaveSnack$^{0.333}$ | WatchTV$^{0.333}$
(2) Reading → Walking R2_Standing_Bookshelf $^{1.0}$

(3) Walking →
R2_Standing Walking $^{0.5}$ | R2 Standing $^{0.5}$
(4) R2_Standing_BookShelf →
R2 Standing Bookshelf Walking$^{0.5}$ |
R2 Standing Bookshelf R2_Sitting$^{0.5}$
(5) R2_Walking → R2_Standing R2_Walking $^{0.5}$ |
R2 _Sitting $^{0.5}$
(6) R2_Sitting → R2 Sitting Armchair $^{1.0}$
(7) R2 Standing → R2 Standing $^{1.0}$

Start.duration = Reading.duration
Reading.duration =
Walking.duration + R2_Standing_Bookshelf.duration
Walking.duration =
R2_Standing.duration + Walking.duration | 1
R2_Standing_BookShelf.duration =
R2_Walking.duration | R2_Sitting.duration

R2_Walking.duration =
R2_Walking.duration + 1 | R2_Sitting.duration
R2_Sitting.duration = 1
R2 Standing.duration = 1

Figure 5. The stochastic context-free grammar with attributes for emergency detection in human activity recognition

- Rule (2) means that the reading activity is composed by two sub-activities: walking in the room in zone R2 (Walking) and standing in zone R2 near the bookshelf (R2_Standing_Bookshelf). The duration for the reading activity will be the sum of the duration of the two component sub-activities.
- The walking sub-activity (the rule (3)) is composed by a sequence of pairs (context: R2, human pose: standing). Also the duration of the walking sub-activity is the sum of the duration of the component sub-activities. The duration of a sub-activity composed by context R2 and human pose standing is 1 unit (the same in case of rule (7)).
- The rule (4) describes the standing near the bookshelf sub-activity, which means taking a book from the bookshelf (context: R2, near the bookshelf and the human pose: standing) followed by the one of the two sub-activities: walking (R2_Walking) or sitting in the armchair (R2_Sitting). The duration of this sub-activity is the duration of walking or sitting sub-activity.
- The walking sub-activity used for R2_Standing_Bookshelf (the rule (5)) is a little different of the walking sub-activity from rule (3). The rule R2_Walking is composed by a sequence of walking sub-activities, which must end with the sitting in the armchair (rule (6)). The duration for rule (6) is considered 1 unit as in case of the rule (7). Each production rule has associated a probability. The sum of the production probabilities for each non-terminal is one.

The context ontology together with the daily programme ontology is integrated in the same ontology, named Home Medical HealthCare Ontology. Figure 6 presents a part of the ontology: (i) The context ontology (describes the living room presented in Figure 2 with 3 zones and some furniture objects); (ii) The daily programme ontology (the properties for several activities: walking, reading, resting and having a snack for a daily programme of a supervised person).

For the moment all the emergency situations have the same level of alert. The model does not distinguish between a low and a high emergency situation.

The parsing process consists of the bottom-up construction of the parse tree. Each sub-activity (which represents an internal node in the parse tree) will be inferred with the daily programme ontology in order to detect emergencies. Also this ontology will be used together with each recognized activity for verifying if an emergency is produced. In the case of a recognized activity the daily programme ontology is used together with the time of the day for verifying if an emergency appears. The context is inferred with each human pose in order to detect a

 ISBN: 978-1-61208-170-0

corresponding emergency, too. If a long duration for the activity/sub-activity, a wrong context, an unusual time of the day for the detected activity/sub-activity or activities conducted in unnatural model are detected, a corresponding message will be generated.

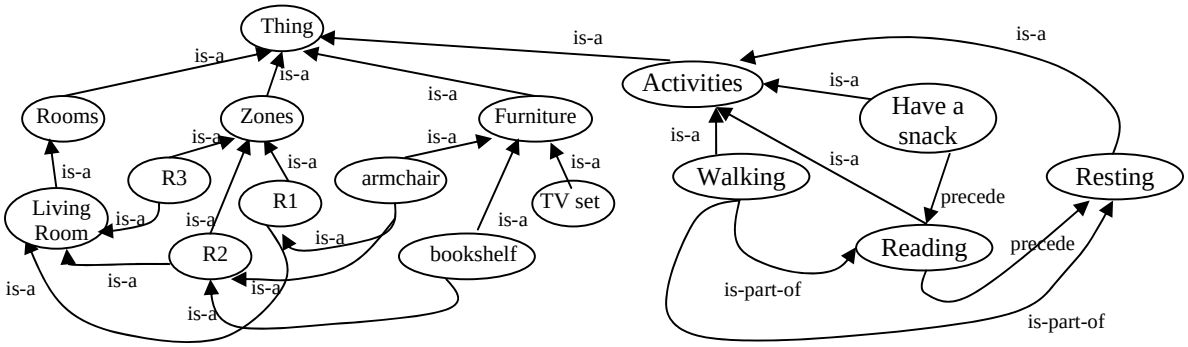

Figure 6. A part of the home medical HealthCare ontology

For example, we consider a set of observations: R2 Standing R2 Standing R2 Standing R2 Standing R2 Standing Bookshelf R2 Sitting Armchair. For this situation the bottom up parsing tree is represented in Figure 7.

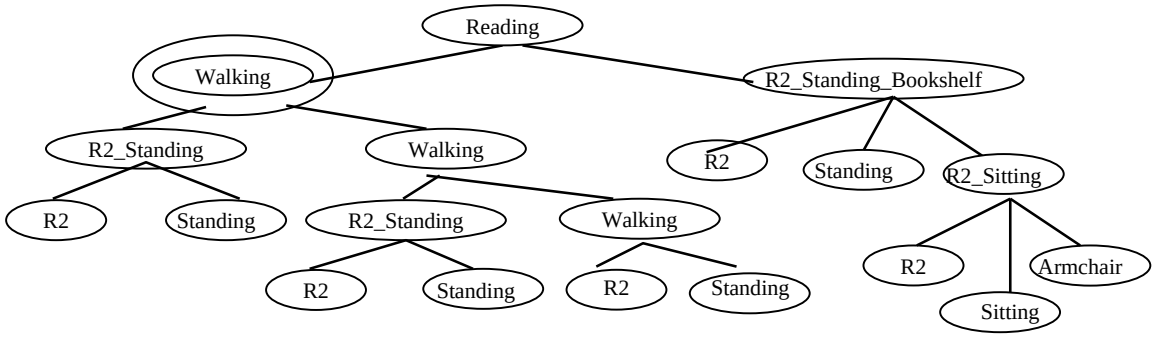

Figure 7. Example of a parsing tree for activity recognition

The emergency detection is checked in the obtained parsing tree:

- Internal nodes whose children are leaves in the tree are first checked by the relation between context and pose; this verification is made in the daily programme ontology based on the location of the sub-activity corresponding to the internal node; in case of a mismatch a context emergency is generated. For example we have a context emergency if there is a node with its children R2 Lying (lying on the floor in R2 zone). In the example from Figure 7 there is no context emergency.
- Internal nodes are checked for the duration of the corresponding sub-activity or activity. This is made using the daily programme ontology based on the duration of an activity. For example, the double circled node Walking from Figure 7 is verified by its corresponding duration. If its duration is longer than a normal value for the Walking sub-activity (the normal duration of the Walking activity is obtained from the daily programme ontology), a duration emergency will be generated.
- The root node is checked by the time of the day in the ontology. If an activity is identified at an unusual time an unusual emergency will be generated. Also the activity is checked by dependent activities. If an unusual order of the activities is detected, an unnatural order emergency will be generated.

V. TESTING THE MODEL

The proposed system was evaluated in a simulated environment. The room is modeled in 3D using 3D modeling software, OpenGL. The supervised human is also modeled using the same tools. Using a 3D modeling tool (the Blender 3D software in this case), we generated the camera and subject 3D models. Then, a few 3D key frames have been created, representing the person's key activities (e.g., walking in the room, standing by the table, sitting by the table, etc). Finally, the intermediate 3D positions have been generated using the interpolating functionality available in the tool. In this simulated environment there is no supervising camera or sensors (depth or movement sensors). In fact, the 3D scene is providing the very same essential information as the image annotation component. The

 ISBN: 978-1-61208-170-0

viewing angle associated with the camera, the depth sensor or with the movement sensor is simulated by the field of view of the viewing frustum. The depth values obtained from the depth sensors are simulated by the distances between each object (furniture object or person) and the position of the camera from the scene. The daily programme ontology is developed in Protégé using OWL. Testing has been performed for the three activities described in Section III (reading, watching TV and having a snack) for which the three emergencies described above: context, duration and time of the day will be simulated. The results have been evaluated using the precision and recall metrics (precision indicates accuracy and recall indicates the robustness). The average values for precision and recall for emergencies detection are about 91.75%, respectively 95.65%, which indicates a good accuracy for the proposed model.

VI. CONCLUSION AND FUTURE WORK

The paper presented a model for daily activity recognition and emergency detection in a smart environment. The recognition process uses a context ontology (for the person's context identification) together with a stochastic context-free grammar (for human pose recognition). A stochastic context- free grammar with an attribute is used for modeling emergencies detection in human activity recognition. The attribute grammar is used for modeling the duration of an activity. The duration of an activity together with the context in which the activity is performed and the time of the day when the activity is produced are used for inferring with the daily programme ontology for detecting emergencies. Based on the promising results obtained in case of this simulated environment, the system will be tested in a real smart environment, in which images from the supervising cameras and a set of collected data from a depth sensor and a movement sensor will be used. We also plan to extend the grammar used for activity recognition to represent a larger set of activities. Also the emergency detection model will be improved with a degree of emergency associated with each such situation. Another development will have as a purpose the recognition of interleaved human activities. In this case the parsing of an activity will be made by identifying a list of sub-trees in the parsing tree from the grammar. The obtained sub-trees will be connected in a single tree by common nodes.

ACKNOWLEDGMENT

The work has been co-founded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the financial Agreement POSDRU/89/1.5/S/62557.

REFERENCES

[1] L. Chen and C. Nugent, "Ontology-based Activity Recognition in Intelligent Pervasive Environments," IJWIS, vol. 5 no. 4, 2009, pp. 410-430.

[2] L. Fiore, D. Fehr, R. Bodor, A. Drenner, G. Somasundaram, and G. Papanikolopoulos, "Multi-Camera Human Activity Monitoring," Journal of Intelligent and Robotic Systems, vol. 52, no. 1, 2008, pp. 5-43.

[3] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors,", IEEE Transactions on Information Technology in Biomedicine, vol. 10, no. 1, 2006, pp. 119-128.

[4] J. Yamato, J. Ohaya, and K. Ishii, " Recognizing human action in time-sequential images using hidden markov model," Computer Vision and Pattern Recognition, 1992, pp. 379-385.

[5] D. Zhang, D. Perez, and I. McCowan, "Semi-supervised adapted hmms for unusual event detection," Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 611-618.

[6] C. Vogler and D. Metaxas, "A framework for recognizing the simultaneous aspects of American sigh language," Computer Vision and Image Understanding, vol. 81, no. 3, 2001, pp. 358-384.

[7] H. Iwaki, G. Srivastava, A. Kosaka, J. Park, and A. Kah, "A Novel Evidence Accumulation Framework for Robust Multi-Camera Person Detection," ACM/IEEE International Conference on Distributed Smart Cameras, vol. 108, 2008, pp. 117-124.

[8] D. L. Vail, M. M. Veloso, and J. D. Lafferty, "Conditional random fields for activity recognition," International Conference on Intelligent Robots and Systems, 2007, pp. 3379-3384.

[9] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine Recognition of Human Activities: A Survey," IEEE Transactions on Circuits and System for Video Technology, vol. 18, no. 11, 2008, pp. 1473 – 1488.

[10] Z. Zeng and Q. Ji, "Knowledge Based Activity Recognition with Dynamic Bayesian Network," Springer-Verlag, 2010, pp.532-546.

[11] M. S. Ryoo and J.K. Agarwal, "Recognition of Composite Human Activities through Context-Free Grammar based Representation," IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 1709-1718, doi>10.1109/CVPR.2006.242.

[12] Y. A. Ivanov and A. F. Bobick, "Recognition of Visual Activities and Interactions by Stochastic Parsing, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, 2000, pp. 852-872.

[13] D. Lymberopoulos, A. Barton-Sweeney, and A. Savvides, "An Easy-to-Program Sensor System for Parsing Out Human Activities," IEEE INFOCOM, 2009, pp. 900-908.

[14] S. W. Joo and R. Chellappa, "Recognition of Multi-Object Events Using Attribute Grammars," IEEE International Conference on Image Processing, 2006, pp. 2897-2900.

[15] S. Mocanu, I. Mocanu, S. Anton, and C. Munteanu, "AmIHomCare: a complex ambient intelligent system for home medical assistance," The 10th International Conference on Applied Computer and Applied Computational Science (ACACOS'11), 2011, pp. 181-186.

[16] I. Mocanu and A.-M. Florea, "A Multi-Agent System for Human Activity Recognition in Smart Environments," unpublished.

[17] I. Mocanu, "From content-based image retrieval by shape to image annotation", Advances in Electrical and Computer Engineering, vol. 10, no. 4, 2010, pp. 49-56, doi 10.4316/AECE.2010.04008.

[18] P. Lyons, A. T. Cong, H. J. Steinhauer, S. Marsland, J. Dietrich, and H. W. Guesgen, „Exploring the responsibilities of single-inhabitant Smart Homes with Use Cases," Journal of Ambient Intelligence and Smart Environments, vol. 2, no. 3, 2011, pp. 211-232, doi 10.3233/AIS-2010-0076.

 ISBN: 978-1-61208-170-0

Environment – Application – Adaptation: a Community Architecture for Ambient Intelligence

Rémi Emonet
Idiap Research Institute
Martigny, Switzerland
remi.emonet@idiap.ch

***Abstract*—This article considers the software problems of reuse, interoperability and evolution in the context of Ambient Intelligence. A novel approach is introduced: the *Environment, Application, Adaptation* (EAA) is streamlined for Ambient Intelligence and is evolved from state of the art methods used in software engineering and architecture. In the proposed approach, *applications* are written by using some abstract functionalities. All *environment* capabilities are exposed as individual services. Bridging the gap between capabilities of the environment and functionalities required by the applications is done by an *adaptation* layer that can be dynamically enriched and controlled by the end user. With an implementation and some examples, the approach is shown to favor development of reusable services and to enable unmodified applications to use originally unknown services.**

***Keywords*-Ambient Intelligence, Architecture, Environment, Application, Adaptation, DCI, SOA, End-User Programming**

I. Introduction

With modern devices and technologies, and with sufficient engineering effort, it is relatively easy to implement smart office and smart home applications. Such applications are usually bound to the considered environment and hard to adapt to a new environment. In the context of Ambient Intelligence, such static application design fails because the user is mobile and the environment evolves continuously. Also, an Ambient Intelligence system is always running and is open: new services (of possibly unknown types) are introduced from time to time. The challenge of software architecture for Ambient Intelligence is to provide a way of maximizing reuse and limiting maintenance. For example, applications should not require any modification or redeployment to handle new service types. Our approach tackles this problem and others.

In this article, we build upon relevant architectural approaches presented in Section II to introduce a new architecture in Section III. We also present, in Section IV, our implementation together with some examples.

II. Approach Foundations: SOA, DCI and others

Our approach can be seen in continuity with previous architectural concepts. In this section, we introduce the architectural concepts that motivate our approach and we provide discussions about other related work.

A. Service Oriented Architectures: SOA

Service Oriented Architectures (SOA) are used in many different contexts ranging from business integration (within and between companies) to Ambient Intelligence. The principle of SOA is to expose software components as "services". Each service encapsulates a particular functionality and provides access to it through a clearly defined interface.

One important characteristic of SOA is "service discovery": a service consumer first queries a service repository (or service resolver) to be able to access a matching provider. Most of the service oriented frameworks work with networked services (a notable exception is OSGi, e.g., in [1]). With networked services, one effect of service discovery is to simplify configuration: service consumers only need to know where to find the service repository.

SOA encourages good encapsulation, loose coupling and abstraction. With little effort, it also helps service consumers adapting to runtime events like the absence or disappearance of a particular service. With encapsulation and discovery, SOA makes it possible to replace a service by another equivalent one, providing the same interface.

As in many other domains, a variety of service oriented initiatives have been proposed but no single standard is clearly dominating. Also, even if service based approaches provide a good way of implementing some "dynamic distributed components", they fail at solving more advanced integration problems. For instance, consider the use case of having an application dynamically (and with no modification) start using services it was not originally designed to use. Such case is typical of an Ambient Intelligence systems where applications and services evolve continuously. Our approach will consider this integration use-case as a common one and not an exception.

The convergence of "Semantic Web" and SOA have been trying to solve the integration problem by letting service designer use their own ontology to describe their services. Ontology alignment methods are then used to make correspondences between services from different providers. Using such correspondence, a service for a given provider can be consumed by a consumer that was designed in ignorance of this particular provider.

 ISBN: 978-1-61208-170-0

In the context of Ambient Intelligence, many projects attempt to integrate different services by building upon both SOA and approaches like the semantic web. Fully automatic service composition and adaptation have been explored, e.g., using multi-agent reasoning as in [2]. Some interesting and well designed approaches are [3] and its evolutions. Also, the soft appliances from [4] envision a systematic decomposition of all existing appliances as independent services. In this vision, end-user programming is used to recreate new innovative appliances from services. One of the main difficulty (and limitation) of end-user programming is to make it both accessible to any end user and powerful enough.

As a conclusion, plain SOA provides a good basis for Ambient Intelligence but it is does not ensure good integration capabilities. We also think that fully automatic approaches are not desired by the end user: these are not optimal and thus can create frustration, and they prevent end users to express their creativity. Classical end-user programming is also too limited to allows at the same time: enabling anyone to customize and innovate with applications, and enabling some users to help in integrating new devices.

B. Data Context Interaction: DCI

Even if it has been studied and practiced since more than 50 years, the domain of software design and engineering is not solved. With time, industrialization methods (waterfall model, UML, etc.) have been created to reduce waste. More recently, "lean" and "agile" methods have taken momentum as they advocate lighter practices and focus on the client.

In our opinion, the most interesting and relevant evolution in recent software architecture and design is the Data Context Interaction (DCI) [5] approach. DCI can be seen as a second attempt to make object orientation (OO) right. The original goal of object oriented programming (and design) was to align the program data model with the user's mental model. This feature is the key to a good human computer interaction: you cannot hide a bad design behind any interface. This becomes more and more important in Ambient Intelligence where user interaction is augmented.

The main principles of DCI are as follows. The *data* objects have the only responsibility to access data (e.g., from a database or memory). In DCI, any use-case of the software is a piece of code that manipulates some *roles*, which are fully abstract. A use-case uses only a set of roles and never manipulates directly data objects. The concept of *role* together with the *context* are the cornerstone of DCI. A *context* is responsible for doing the mapping of some roles onto some concrete data objects. The context is populated in response to user interaction (e.g., selecting things then clicking on a button) and then the use-case is executed using this context.

As an example, we can consider a banking application with the use case of making a money transfer between two accounts. More precisely, consider the MoneyTransfer use-case: it involves three roles that are the SourceAccount role, the DestinationAccount role and the MoneyAmountProvider role. The MoneyTransfer code will start a transaction, then query the amount to transfer from the MoneyAmountProvider, then call $withdraw$ on the SourceAccount and call $credit$ on the DestinationAccount. The context is created and populated by the application when the user is asked to select a source account (e.g., his CheckingAccount data object) and a destination account (e.g., one of his SavingsAccount) and an amount (e.g., could be just a plain "int" value).

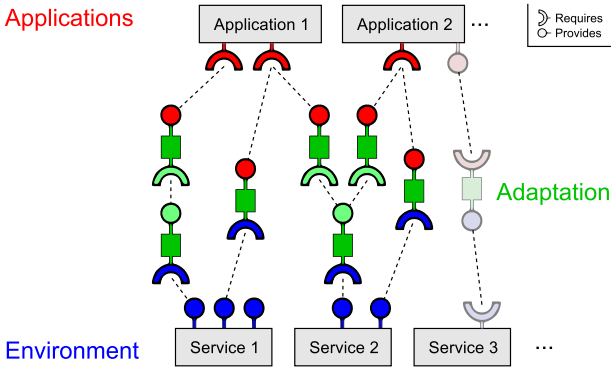

Figure 1. Proposed EAA architecture – *Environment* provides low-level services. *Applications* manipulate only high-level abstract services. *Adaptation* bridges the two and is dynamically extensible and user-controlled. Lighter chain on the right: inversion due to the whiteboard pattern.

C. Other Related Work

A mobile agent is an autonomous programs that can migrate between computers over a network. Even if this is an interesting feature for Ambient Intelligence, it can be seen as orthogonal to the subjects discussed in this article and can complement the proposed approach. An example of using mobile agents as an infrastructure is presented in [6].

The domain of human computer interaction tends to evolve from desktop-like applications to Ambient Intelligence. In this context, an emphasis is put on how to dynamically split and distribute user interfaces based on the available devices. The concept of meta-User Interfaces (meta-UI) has been introduced in [7] and consists in having an interface to control and introspect an Ambient Intelligence environment. A deep and interesting analysis related to our problems is conducted in [7], however, their application is limited to the migration and adaptation of graphical user interfaces between devices.

III. Proposed Approach

In this section, we introduce our Environment, Application, Adaptation (EAA) approach and how it can interact with a community built around it. In the same way as DCI is an attempt to make OO right (see Section II-B), EAA is an attempt to make SOA right.

 ISBN: 978-1-61208-170-0

A. Environment, Application, Adaptation

The Environment, Application, Adaptation (EAA) approach builds on top of Service Oriented Architectures (SOA) and takes similar inspiration as Data, Context, Interaction (DCI). In EAA, most of the elements are services: in some sense, services act as objects (with interfaces) that can be distributed and dynamically discovered. As in SOA, the capabilities of the *environment* are exposed as plain services in EAA. In a parallel with DCI, these environment services are corresponding to the data part from DCI.

Most importantly, EAA has the equivalent of roles in DCI. Any *application* only manipulates some abstract services (roles) that correspond to its exact requirements. The design of the application is done without bothering about what concrete service can or will be used to fulfill the role. With this choice, the environment will never directly provide any service that an application need.

In DCI, the context is responsible for the casting: concrete data objects are recruited to play some roles. In EAA, the *adaptation* layer is responsible for the equivalent, which consists in using services from the environment to create services required by the applications. The adaptation layer is populated through implicit or explicit interaction with the end user (same as in DCI).

In Figure 1 (ignoring the lighter rightmost elements), a set of applications, environment services and adapters are shown. Colors are used to distinguish service types coming from the environment (in blue), the applications (in red) or the adaptation (in green).

B. Using Service Factories for Adapters

To populate the adaptation layer, some adapter factories are used. Each factory is actually a service that exposes which kind of adapters it can create and that creates it on demand. The concept of service factory is taken from [8] and restricted to adapters: we do not consider the case of open factories that can create services without requiring any another service. With our restriction, the number of instantiable adaptation paths becomes finite and it is thus possible to filter and display them to the user (see Section IV-B).

C. Refinement using the Whiteboard pattern

A useful pattern in service oriented design is the "whiteboard" pattern [9]. The goal of this pattern is to simplify the design of clients of a particular service. Let's consider a Text2Speech service that is designed to receive some text sentence and will output it as speech through loud speakers. In a classical approach, any client of the Text2Speech service would first look for the service, then connect to it and then send the message to it. Eventually, the search-and-connect code is here duplicated in all clients.

Using a whiteboard pattern, the situation is reversed and the Text2Speech service is actually doing the search-and-connect. Each client just declares itself as Text2SpeechSource and the Text2Speech will connect to it as soon as it finds it. With the whiteboard pattern, some code is moved from the client to the "server", which limits redundant code writing and makes backward compatible evolutions easier (the server handles the various versions of clients). From a service point of view, now the "server" looks for its clients, which causes an inversion of the provides/requires dependency as shown in Figure 1 (on the right) and in Figure 2 (lower part).

In EAA, the whiteboard pattern is typically used on the view side, i.e., when the application state needs to be brought back to user (through the environment). The above example of voicing the output of an application using a Text2Speech service is a typical example of this.

D. Community Architecture and Sharing

The structure of the proposed EAA makes it a "community architecture" [10] in a double sense. First, the approach encourages the creation of a community around it and provides a structure for it, and second, it is the community itself that is creating the actual, live, evolving architecture.

We distinguish four entry points in EAA for innovation and extension, each requiring different skills. Compared to some end-user programming approach where there is trade-off to make between the expressive power of the programming and the required skills to use it, EAA has multiple values for this trade-off. It would be interesting to investigate how EAA can be combined with an end-user approach targeting more ease of use than power of expression (higher expression power being provided by EAA).

The first two entry points are for a relatively large audience. First, most end users will be able to innovate at the adaptation level by doing a smart and original choice of adapters for a particular application in their environment. Also, any end user can take part in the community by suggesting new ideas for services, applications or adapters. With proper documentations and examples, we can expect a reasonable part of the users (surely less than 10%) to be able to create new adapters by copying an existing one or using a wizard tool (in current implementation, an adapter is just a XML file and thus it is quite easy to define new ones).

More advanced extension points concern the contribution of new applications or new environment services. Both require more advanced computer skills but really different ones. Application developers will probably write their application and maybe a couple of adapters to integrate it into the existing ecosystem: the skills required here are mostly classical application development skills. The contributors of new environment services will probably be people that like hacking with new devices or new signal processing methods (image or audio processing, accelerometers, etc.): their goal would be to innovate by providing innovative input or output medium to transform existing application.

 ISBN: 978-1-61208-170-0

The EAA does not define by itself what kind of services are used by the people. It is the community itself, by creating new environment services, applications and adapters, that decides on what is the actual architecture. We cannot rely on any user to make the best architectural choices. However, if the community is sufficiently large and open, we can expect to find a small proportion of "architects/moderators" as in other open community projects: their role could be for example to avoid proliferation of totally similar concepts and avoid fragmentation of the community.

IV. Implementation and Example

To experiment with the proposed approach, we implemented different test cases. In this section, we provide some implementation details and explain these test cases. More details can be found with the source code that will be provided online, see http://its.heeere.com/ambient2011 .

A. Implementation Details

We implemented the whole presented approach letting aside only the community aspect (e.g., dedicated system for sharing adapters). Our implementation is based on the open-source OMiSCID [11] service-oriented middleware. We created a set of small reusable services and designed a graphical user interface for the user to control the environment and the adapters. The applications are implemented as services that explicitly require some functionalities. Functionalities from the environment are exposed as OMiSCID services.

The developed services will be made available online and include the following services: exporting a display area (on a screen or video project), exporting a mouse pointer, and exporting a "chat" service to allow to open popup messages on a computer. Also, under Linux operating systems, we provide additional features such as a text-to-speech service based on "espeak" and a service to generate synthetic keyboard events on a computer (this one is used for example to control presentations or games).

For the adapters, we designed a generic program that takes an XML description of a family of adapter and starts the corresponding adapter factory (that can start an adapter instance on demand). The XML description contains information about the adapter such as which functionality it takes as "input" and to which one it converts it. The adaptation code, that is usually simple, can be provided within the XML file using ad'hoc languages such as JavaScript or XSLT.

With the assistance of a graphical user interface, the final user can decide what adapters to eventually use. The end of the next subsection is dedicated to the illustration of the simple graphical interface we implemented to help the user managing the environment and the adapters.

B. Detailed Test Case

To showcase our approach, we detail the case of a simple tic-tac-toe game we developed. For now, we consider that

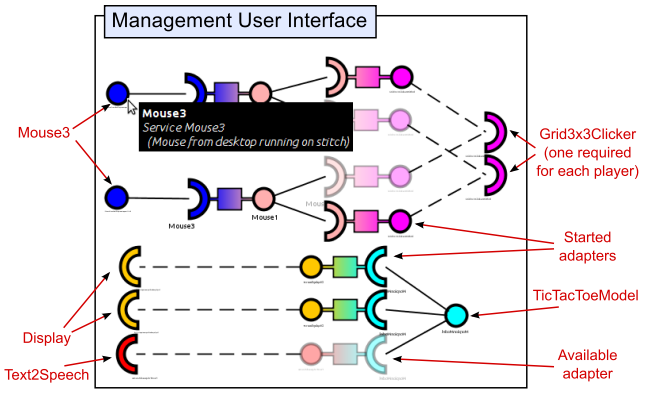

Figure 2. User interface to manage adapters (annotated screen capture, best viewed in color). The panel shows all required services, available services and possible adapters obtained from factories using service discovery. A color code is associated to each service type. By clicking on a instantiable adapter (lightened ones) the user can ask the system to create this adapter. Note that due to the whiteboard pattern, the provides/requires relation is reversed for the display side (lower half of the panel) where the environment requires services from the applications.

the environment contains only two computers, and from each one we exported some services: a Display, a Mouse3 (mouse pointer with 3 buttons) and a Text2Speech. Each exported Display service has a unique identifier and follows a whiteboard pattern to connect to any matching DisplaySource it finds. A DisplaySource is expected to send drawing commands to the Display. The game logic is implemented as a service that exposes a TicTacToeModel and that follows a whiteboard pattern for two services: two Grid3x3Clicker with two different unique identifiers.

To bridge the gap between the environment (Display, Mouse3) and the application (Grid3x3Clicker, TicTacToeModel), we introduced a set of simple adapters. The first ones are for input and can be heavily reused in other context: one adapter converts a three button mouse Mouse3 to a single button mouse Mouse1, the second adapter converts a Mouse1 to a Grid3x3Clicker by converting click x, y position to some grid index from 0 to 8. We could have skipped the distinction between Mouse3 and Mouse1 but we kept it as it is useful is some other contexts. On the display side, a specific adapter was written to convert TicTacToeModel to a DisplaySource: the tic-tac-toe state change events are converted to drawing commands such as drawing circles.

From the list of functionalities required by applications and functionalities exposed by the environment, the user interface considers all available adapter factories (also exposed as services, see Section III-B) and proposes possible adaptation paths. Figure 2 illustrates the user interface we designed to allow the user to manipulate adapters. The user sees all possible adaptation paths, and an instantiable adapter is clicked, the system automatically queries the corresponding adapter factory to create the adapter.

 ISBN: 978-1-61208-170-0

By letting the user control the adaptation layer, EAA makes the tic-tac-toe become ambient. The use of properly decoupled services (SOA done right) makes it possible for the user to dynamically select where and how to display the game and how to control it. EAA, with its explicit adaptation layer, makes it also possible to easily create variations of the game that integrates into an Ambient Intelligence vision. To this end, different adapters can be used. A first adapter, which is simple but specific, transforms the game state (TicTacToeModel) to some short textual output to be processed by a Text2Speech service. A reusable adapter, used for input of the game, uses a SpeechRecognizer and converts voice commands such as "play in three" to a Grid3x3Clicker. In addition to the audio modality, computer vision is also used as a possible input: by sticking post-its on a surface, the user can transform it to a Grid3x3Clicker thanks to a dedicated adapter.

C. Other Test Cases

Apart from the tic-tac-toe, we also implemented other environment services, games, applications and adapters. For example we created a MagicSnake game that consists in guiding a snake in a 2D maze to reach a target as fast as possible while avoiding walls. As an experiment, we also modified a game called "Nuncabola" where the player controls a ball rolling in a 3d environment. Both games use a two dimensional analog input: we implemented these input with different combinations of environment services and adapters. Eventually, we control these games using:

- obvious device such as a mouse or a keyboard,
- more exotic devices such a accelerometer-based devices (e.g., smart phone, WiiMote) or WiiFit-like devices,
- computer vision and human tracking (e.g., the player moves in the room to control the ball acceleration, or the player moves his hands, arms, etc.)

Using simple generation of keyboard events, we also implemented a slide presentation controller. We used various methods to skip to the next/previous slide including for example computer vision, e.g., gestures; sound recognition (clapping hands); and voice recognition, e.g., saying "next".

V. Conclusion and Future Work

This article presented the Environment, Application, Adaptation (EAA) architectural approach. With this approach, the environment and the applications are fully independent of each others. This both encourages the design of more generic environment services and eases the deployment of an unmodified application in a new environment: this deployment is possible even if eventually the application ends up using only originally unknown services. The glue between what a particular environment offers and what a particular application requires is done by a dedicated adaptation layer. This layer makes the overall system easier to adapt and open to user control and innovation.

An implementation of this approach was showcased: this implementation is fully operational and allows dynamic runtime extension with new services, applications and adapters.

The main future directions involves the improvement of the user interface (icons for service types, quick filtering, etc), and the setup of the community infrastructure to make it easier for users to use and contribute innovative adapters.

Acknowledgment

The author would like to thank the PRIMA research group and particularly Matthieu Langet for his tight collaboration.

References

[1] C. Escoffier and R. Hall, "Dynamically adaptable applications with iPOJO service components," in *Software Composition*, 2007, pp. 113–128.

[2] M. Vallée, F. Ramparany, and L. Vercouter, "Dynamic service composition in ambient intelligence environments: a multi-agent approach," in *Proceeding of the First European Young Researcher Workshop on Service-Oriented Computing*, Leicester, UK, April 2005.

[3] M. Assad, D. Carmichael, J. Kay, and B. Kummerfeld, "PersonisAD: distributed, active, scrutable model framework for context-aware services," *Pervasive Computing*, pp. 55–72, 2007.

[4] J. Chin, V. Callaghan, and G. Clarke, "Soft-appliances: A vision for user created networked appliances in digital homes," *Journal of Ambient Intelligence and Smart Environments*, pp. 69–75, 2009.

[5] J. O. Coplien and G. Bjørnvig, *Lean Architecture: for Agile Software Development*. Wiley, 2010.

[6] R. Razavi, K. Mechitov, G. Agha, and J. Perrot, "Ambiance: a mobile agent platform for end-user programmable ambient systems," in *Proceeding of the 2007 conference on Advances in Ambient Intelligence*. IOS Press, 2007, pp. 81–106.

[7] J. Coutaz, "Meta-user interfaces for ambient spaces," *Task Models and Diagrams for Users Interface Design*, pp. 1–15, 2007.

[8] R. Emonet and D. Vaufreydaz, "Usable developer-oriented functionality composition language (ufcl): a proposal for semantic description and dynamic composition of services and service factories," in *Intelligent Environments, 2008 IET 4th International Conference on*. IET, 2008, pp. 1–8.

[9] O. Alliance, "Listener Pattern Considered Harmful: The Whiteboard Pattern, 2nd rev." http://www.osgi.org/documents/osgi_technology/whiteboard.pdf, 2004, [Online; accessed 28-July-2011].

[10] F. Moatasim, "Practice of community architecture: A case study of zone of opportunity housing co-operative," Ph.D. dissertation, McGill University, 2005.

[11] R. Emonet, D. Vaufreydaz, P. Reignier, and J. Letessier, "O3miscid: an object oriented opensource middleware for service connection, introspection and discovery," in *1st IEEE International Workshop on Services Integration in Pervasive Environments*, 2006.

 ISBN: 978-1-61208-170-0

The Experience Cylinder, an immersive interactive platform

The Sea Stallion's voyage: a case study

Troels Andreasen, John P. Gallagher,
Roskilde University
CBIT, Building 43.2
Universitetsvej 1
Roskilde Denmark
{troels,jpg}@ruc.dk

Nikolaj Møbius, Nicolas Padfield
Roskilde University / illutron
HUMTEK, Building 08.1
Universitetsvej 1
Roskilde Denmark
{nimo,nicolasp}@ruc.dk

***Abstract*—This paper describes the development of an experimental interactive installation, a so-called "experience cylinder", intended as a travelogue and developed specifically to provide a narrative about the Viking ship Sea Stallion's voyage from Roskilde to Dublin and back. The installation provides a framework for individual experience of the Sea Stallion's voyage, where a user among the audience interacts with the installation. Interaction is mainly by position and gesture tracking and feedback is by selection and projection of objects, such as photos, videos, background images, synthesized sound and recited stories. All combine to provide an individual path of choice through a coherent story told by means of a cylindrical screen, enclosing a space providing 3D positional sound following the interacting user and exploiting data such as wind, wave and weather logs through 3D sound effects. While being intended to tell a story about a certain voyage the resulting installation has been developed as an open interactive platform suitable for story-telling/travelogue. First and foremost the platform provides an approach to easy and intuitive navigation within heterogeneous media material comprising a large number of media objects.**

Keywords - interactive installation, bodily navigation, cylindrical screen

I. INTRODUCTION AND MOTIVATION

The "experience cylinder" described in this paper is an interactive installation developed in connection with a pilot project carried out in a newly established Experience Lab at Roskilde University [1]. In collaboration with the Danish Viking Ship Museum the aim was to establish a "story-telling" platform that presents the documentation of the reconstructed Viking ship Sea Stallion's (Havhingsten) voyage from Denmark to Ireland and back in 2007-2008 and provides an individual experience of the voyage to the interacting user.

The physical surroundings comprise a continuous image on a large cylindrical screen (six meters in diameter) provided by several connected projectors, ambient positional sound provided by several speakers and precision position as well as gesture tracking using 3D cameras. The story of the voyage can be told in this environment individually to users walking around, approaching what captures their interest. The design of the cylinder allows users to obtain more details through images (photos, video, animations) and sound (video soundtracks, recitation of logbook and stories, wind and wave data reflected by artificial sound effects) that emerge as a response to movement and can be further invoked by gestures.

In its present incarnation, the focus is on presenting historical knowledge and an engaging narrative in a museum context. Our and the museum's ambition is to use recent technology for the interactive presentation of historical material in a way that does not resemble the PC-like 'kiosks' with touch screens often found in museums. The aim is to give the user of the cylinder a sense of presence and of personal exploration in the story of the Sea Stallion's voyage, as opposed to a passive experience as recipient of a predetermined narrative.

The installation comprises, however, an open platform facilitating navigation through any amount of heterogeneous media, such as video, audio, text, photography, nautical charts and weather data. We are experimenting with new embodied input models where audience movement and gestures are interpreted as expressions of interest and used to determine focus and navigation.

We are in the midst of an iterative process; first, two mockups were programmed on a standard computer using a normal screen and mouse for input. Then the full scale installation was built and the software tested. Some changes were made and the installation presented and tested at a workshop. The next stage includes more extensive and formalized user testing, more data to be displayed, and in time different data sets.

II. THE EXPERIENCE CYLINDER

The experience cylinder, a section of which is shown in Fig. 1, consists of a 6 meter diameter projection screen. Participants enter this 360 degree other-worldly immersive experience by sliding past a section of curtain, the curtain serving both as the projection screen and as the visible boundary to the technologically enhanced reality within the cylinder, where normal physical laws are supplemented with sensors allowing control of the surroundings just by moving about.

 ISBN: 978-1-61208-170-0

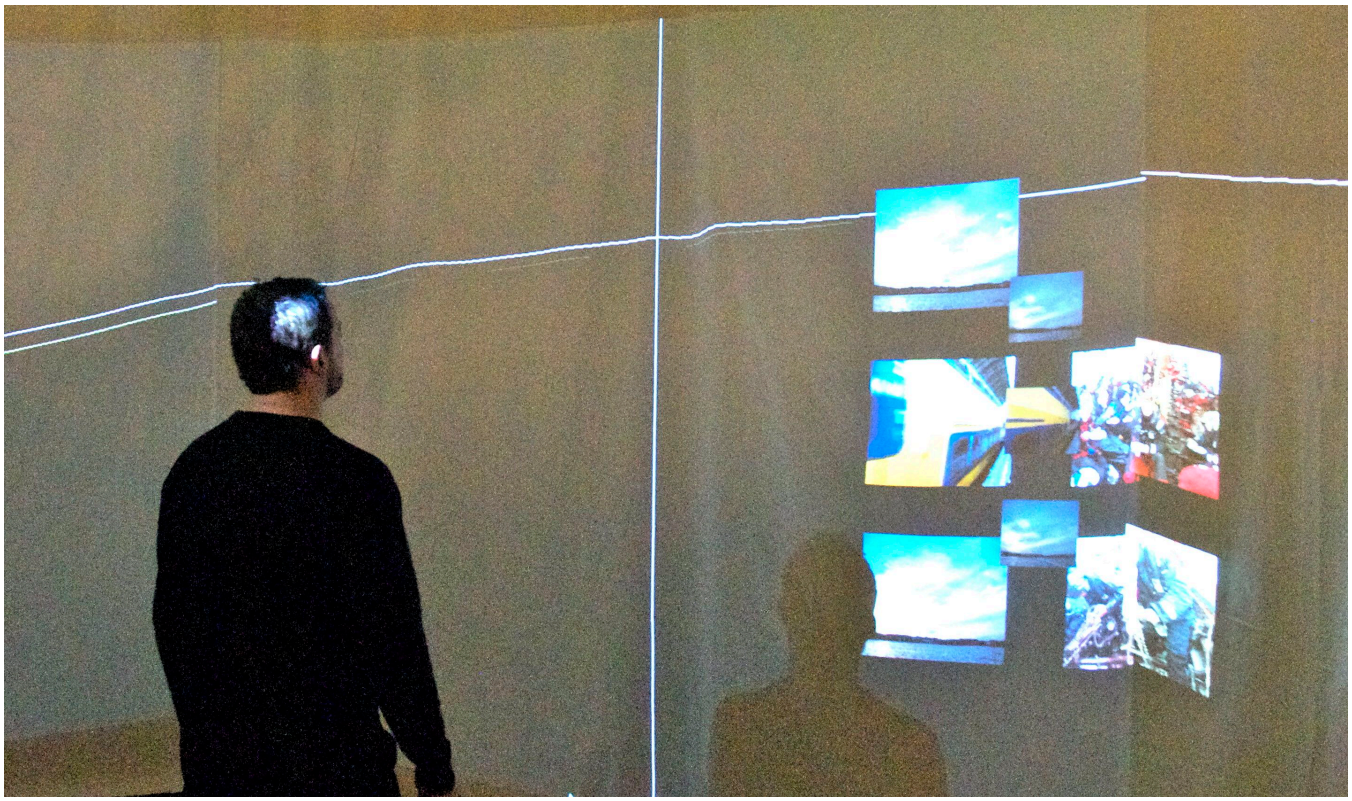

Figure 1. The experience cylinder.

Media are displayed all around the inside wall of the cylinder, reminiscent of pictures hanging on the wall at a gallery. The initial state is one of rest, all media about the same size, no video or sound playing, only a calm sea as background visual and ambient sound. When the participant signifies interest by moving towards a cluster of media, those media are fluidly expanded, more details and more media objects appear and video or sound clips can be activated. The participant can thus by merely moving around navigate a very large media collection and zoom in upon any particular single element, without use of any traditional computer UI elements and metaphors such as windows, cursors or mice.

A. System design

A required property of the design was from the beginning the ability to navigate through the trip following the route with a possibility to walk or jump to certain sections of interest to immerse into more detail. So a simple first solution was to draw a map of the route on the floor according to the geo-coordinates of the track, a particular point representing geographical as well as temporal location in the narrative as illustrated in Fig. 2, and to support navigation by walking this map.

While such a design would have fulfilled the main purpose of presenting the Sea Stallion's voyage, we moved away from this as we felt it would not function optimally as an engaging installation, being very predictable and lacking the ability to inspire participants "by surprise" – and thereby stimulating participants to experiment with the interaction in a playful manner.

Having discussed and compared several solutions for navigation input, including a variation of simplifying projections of the track image, we finally agreed on an alternative that also implicated a model interactive framework as a whole – a circular projection as a metaphor for the route track enclosing the interior disk as a navigation space, as illustrated in Fig. 3. The obvious addition to this was a cylindrical projection screen leading to the basic design of the experience cylinder.

Thus in this setting the circular extent represents the track as well as a timeline for the voyage. Media are organized on the screen corresponding to temporal placement in the duration of the voyage. The ship's departure from Roskilde is at "0 degrees", the middle of the voyage, landfall in Dublin, is at "180 degrees" and arrival at home port of Roskilde again is at "360 degrees". This arrangement seemed intuitively graspable by most participants – seeing the floor of the circular space as a large clock face laid flat, where further clockwise corresponds to 'later'.

The experience starts in the middle of the circular area. The entire journey is visualized divided into an appropriate number of pieces shown on a corresponding number of equal sized circular segments on the canvas as indicated in Fig. 3. By turning around in the middle, one can assess the entire trip. Immersion into more detail of a whatever part of the trip that captures ones interest can be done simply by moving in the direction of this part.

 ISBN: 978-1-61208-170-0

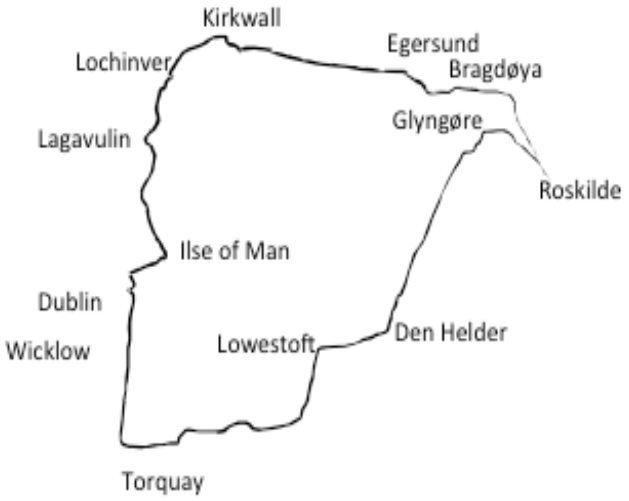

Figure 2. Geo-coordinates of the track.

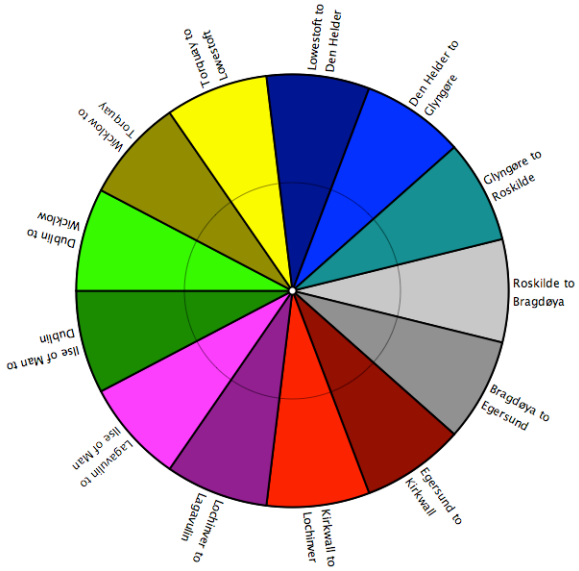

Figure 3. A circular projection enclosing a navigation space.

This causes the map to be “re-projected" so that the segment in focus grows and other segments shrink. This magnification is progressive - the distance from the center determines the degree of magnification. Fig. 4 shows an example of a projection based on the indicated position on the disk – the user has moved towards the “Den Helder – Glyngøre” segment to study this in more detail. When the distance to the center exceeds the appropriate limit (here the half-radius) then segment division will be frozen allowing the viewer to walk around inside the corresponding pie to study details of part of the trip in focus without being disturbed by further re-projections. By moving back towards the center and continuing in another direction, a new part of the trip can be investigated. Or by going in circles around the center, the whole journey can be traversed in more or less detail depending on the viewer’s distance to the center.

Thus the interaction results in feedback in the form of a modified screen division and an adapted presentation of media objects. Each point on the circle represents a position on the trip (a time slot). Each pie represents a piece of the trip and the period for which this was completed. The selection of media objects adapts such that more objects and more details are presented on large pies and less details are shown on small.

Media objects are as mentioned presented by projection on the cylindrical screen and by means of the 3D positional sound. Objects that appear and float as projections on the segments of the screen include segment titles, photos, videos, stories (shown as titled icons) such as fragments from the log or tales on themes such as seasickness. Video and story objects involve sound and occupy time spans and therefore activation and deactivation options for each of these are needed in the interaction. These media objects can be activated by arm pointing gestures (or by approaching close enough) and deactivated by moving away or activating other objects.

Obviously the zooming functionality explained above in itself will enable set-ups with a significant amount of media objects (the Sea Stallion voyage has about 25000 photos). However, to improve support of huge amounts we are experimenting with ways to include 3D rendering of 2D media with objects positioned at different depths in a 3D space projected on the screen. Thereby a new means and purpose of interaction arise: a user can look behind an object covering another by moving to the side. This 3D-rendering is illustrated in Fig. 5. The challenge here is to intuitively combine the basic zooming feedback with 3D perspective behavior.

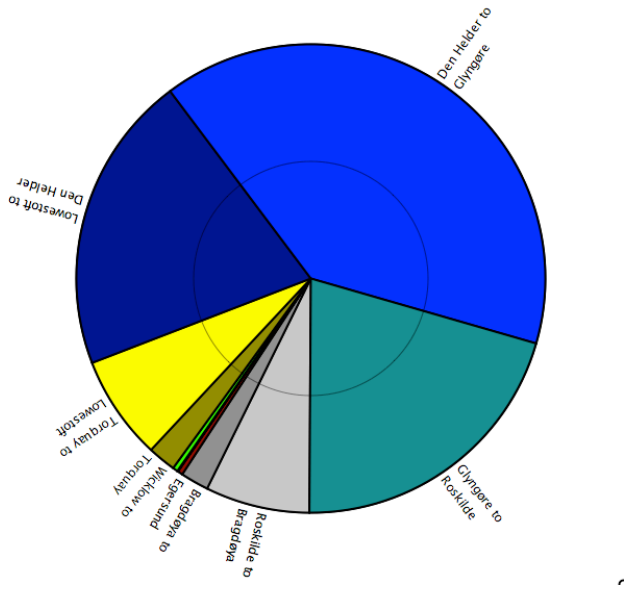

Figure 4. Interacting by walking – moving towards a section of interest.

In addition to the primary objects mentioned above the installation also incorporates a seascape animation by means of a visual as well as auditory background to provide the visitors with a feeling of being on a ship at sea. The background visual rendering exploits the 360 degree sea level "horizon" animating waves as well as rolling of the ship. The visual horizon is accompanied by a synthesized background sound imitating wind and rolling waves.

This seascape animation can also be used to exploit data from the voyage. A significant amount of detailed data about waves and weather conditions logged on geo-positions is used to reflect the weather conditions and intensity of the waves at the time and position of the media currently being focused on.

III. IMPLEMENTATION

A. Tracking

Tracking the participants is accomplished with a Microsoft Kinect® 3D webcam and computer vision algorithms. The Kinect sensor functions by projecting a special pattern with an IR (infrared) laser, which enables the IR camera to calculate the distance to all objects in the field of view by measuring distortion of the known pattern. The Kinect delivers distance data in the form of a greyscale value for each pixel. Until recently, tracking individuals would have required IR floodlighting the whole area and applying background subtraction, frame differencing and finally blob detection to images captured by an IR security camera. The Kinect now makes it far easier to recognise a person, simply using blob detection.

The system is programmed to lock onto the participant in the center of the circle and follow this individual until she leaves the field of view. This means any number of people can be in the installation at one time, without confusing the system, the interaction or each other. The only challenge with the Kinect is engineering a large enough field of view, as it (being a consumer device) does not support interchangeable lenses. The easiest is to have a high ceiling, or, failing that, to stitch together multiple images in software.

B. Software architecture

The installation is programmed in OpenFrameworks [10], an open source C++ toolkit for creative coding. OpenFrameworks provides easy access to multiple libraries we require (OpenGl, Quicktime and freeType among others) and can run on Mac, Windows and Linux.

The sound is created and controlled in MaxMSP [8]. The positioning of sounds etc. is controlled by MAX, which receives commands from OpenFrameworks via UDP, enabling the work to be distributed to several computers.

The visualization is written to a virtual buffer in OpenFrameworks, offering the possibility of distributing work across several computers and in the future supporting projector overlap with alpha blend for a more seamless display, higher light intensity and easier adjustment.

.

Figure 5. 3D display of media objects.

The seascape background rendering is based on Virtual Choreographer [5] for the interaction (message capture and interpretation) and on OpenGL for the geometrical rendering. The output frame can be captured in a Frame Buffer Object and distributed over different viewports to cope with the distribution of the graphical rendering over 3 GPUs.

C. Physical construction

For the physical construction of the installation we utilised rapid prototyping principles to quickly get to a testable stage. We were interested in using as many standard off-the-shelf elements as possible both to keep costs down and to increase flexibility, but did have to custom produce certain elements.

The construction is illustrated in Fig. 6 and 7. The 6 projectors (an Acer widescreen 4000 ANSI lumen model selected for its good price/resolution/light output ratio) are driven by one Mac Pro computer with 8 cores and 2 graphics cards, via 2 Matrox Triple Head 2 Go, which split the computer's 2 DVI outputs to 6 VGA outputs (VGA being selected as it is easier to run long cables).

A 7 channel surround sound system consisting of Prodipe monitor speakers mounted at head height, supplemented by 4 subwoofers, is driven by an 8 channel M-audio Firewire sound card connected to the Mac Pro.

The projection screen is standard cheap unbleached cotton cloth.

We had to custom weld a modular circular lighting rig of 50mm standard tubing to carry the circular projection screen and the projectors. After trying out different commercially available projector mounts, we found it easiest and cheapest to produce our own, CNC milling them out of 12mm plywood. The whole rig is suspended on 3 chain hoists enabling it to be raised and lowered for easier adjustment.

 ISBN: 978-1-61208-170-0

Figure 6. Construction of the experience cylinder.

D. Media and narrative

The material made available by the Viking Ship Museum for the project involves mainly the two parts of the journey:Wicklow to Torquay and Den Helder to Glyngøre. The material includes photos, video, position, weather and wave data recorded at locations along the route. In addition we have had access to selections of texts, including log books and descriptions related to a theme and stories written by crew members.

All photos and video footage taken by photographer Werner Karrasch from the Viking Ship Museum. These are unique and impressive recordings that are ideally suited to "dramatize" the tour. Much of the footage is taken from a dinghy that followed the Sea Stallion at a distance and managed to give a very vivid impression of movements on the high seas.

The installation does not include any direct display of text apart from simple words and titles for parts of the trip. However, the installation includes functionality for playback of audio clips, and texts such as logs and stories related to certain themes, which are recited and included as clips.

Weather and wave data are presented by combining synthesized wave, wind and rain sound and artificial sea level animation as background images. Currently sound modeling as well as background animation is developed, but this is not coupled to the data. Some work has yet to be done on data preparation and modelling.

IV. CONCLUSION, PERSPECTIVES AND FURTHER WORK

A. Evaluation and Related Work

The main result and contribution of the project is to create a physically interactive environment - the experience cylinder - in which precise tracking of body movements results in immediate responses in the visual and audio display of media items on the cylinder's screen and speakers. The cylinder has been applied to an experiment in which the items displayed relate to the voyage of the reconstructed 11th century Viking ship Sea Stallion from Denmark to Ireland and back. Initial evaluation in two workshops with about 25 participants each gave valuable user feedback and suggestions for further work. Firstly, experience shows that the cylinder gives the user a sense of control and engagement in that the display reacts smoothly and coherently to personal movements. Secondly, the cylinder can “tell a story" even though the user can in principle move freely in the space, the physical arrangement of the media items providing a certain narrative timeline structure. The cylinder has a domain-independent architecture, in that the media content relating to the Sea Stallion is independent of the tracking and display control, and thus the cylinder could be adapted for other applications.

The ideas behind the experience cylinder evolve from earlier concepts such as “physically interactive environments” [9], [11], “immersive virtual environments” [13], and “tangible interfaces" [12], originate in Kruger’s seminal ideas dating back to the 1980s on computer-controlled interactive spaces [7]. Our system uses low-cost and easily available technology, as do other recent systems such as the CryVe system [6]. However, the system described here differs from other typical virtual and mixed-reality systems in that it aims to allow a user to navigate using physical movements though a mass of heterogenous but interrelated media, rather than place the user in some specific virtual spaces.

Numerous examples appear in the literature of the development of interactive environments for the general purpose of communication, in areas such as advertising, entertainment, story-telling and dissemination of cultural assets; see for example [2], [4], [11]. One of the main motivations for such developments seems to be that physically interactive environments are perceived to offer the user a greater sense of presence and immersion, allowing the user to engage more actively with the content of the communication than is the case with traditional media. This is particularly so when interaction involves substantial physical body movements and gestures; Falk and Dierking [3] argue that the human sense of self is closely bound to physical interaction and that this is vital when considering individuals’ learning experience in the museum context. Our experiment shares the same general motivations, and exploits newly available affordable 3D camera technology for precise body tracking, allowing refined physical interaction such that small movements provide immediate and precise visual and audio feedback. We conjecture that the new concept of “tangible” computing is not necessarily related to tangible objects alone, but that new technology enables spatial interfaces which allow us to leverage humans’ well developed spatial reasoning abilities in UI and invent new metaphors for engagement.

The concept of narrative is vital to the experience cylinder. In the case of the Sea Stallion's voyage, the physical structure of the circular screen in itself imposes a narrative structure corresponding to the route from Roskilde to Dublin and back, even though the user can "follow" the

 ISBN: 978-1-61208-170-0

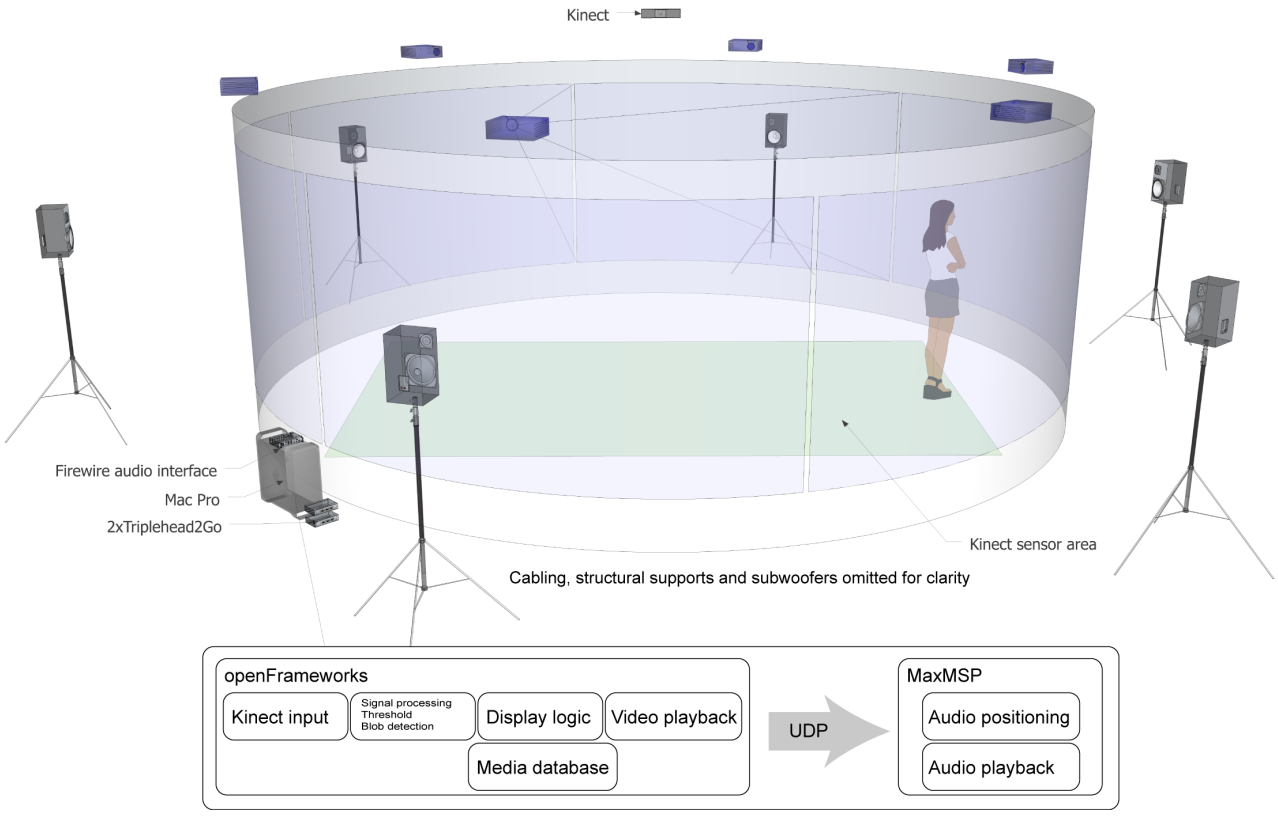

Figure 7. Diagram showing layout and main hardware and software components.

narrative in any desired order by moving around in the space. By contrast, the temporal order, size and arrangement in which media items appear on the screen can be controlled by the designer, allowing local storylines to be imposed within the overall structure. These aspects are the subject of continuing experiment and evaluation in the experience cylinder. The concept of "narrative spaces" is further explored by Sparacino [14], who offers promising narrative models, merging ideas from artificial intelligence such as probabilistic interaction models with physically interactive environments.

B. Future Work

Preliminary experiments and demonstrations confirmed that the cylinder has considerable potential to deliver an immersive, engaging experience that can be personalized and that the platform can combine knowledge transfer and story-telling with the feeling of being present on the journey. At the same time, the experience cylinder showed clearly the need for much further research in both technical and communicative dimensions. Firstly, the platform itself can be improved by technical refinements and extensions, including the addition of more 3D sensors covering the area with greater precision; this will also contribute to extensions allowing gesture tracking and tracking several users simultaneously. The quality of the projection and rendering can be improved with, for example, back projection, while alpha blending allows transparency effects and a more seamless display. Secondly, the narrative elements and structure will undergo further investigation, experiment and evaluation. Questions that arise include how to provide better visual and audio orientation points that maintain the narrative structure, and how context such as the user's previous route or prior knowledge could influence the interaction. Thirdly, the Sea Stallion voyage generated a huge amount of data, which motivates the integration of techniques for intelligent searching, for example based on an ontology relating to the voyage and its historical background. The intention is that the user can explore a chosen theme within the huge amount of available media, such as for example life on board the ship, shipbuilding methods or Viking history.

ACKNOWLEDGMENT

The project was partially funded by RUCInnovation, Roskilde University. We would like to thank the Viking Ship Museum, Roskilde, for supplying the media relating to the Sea Stallion's voyage and for their support and enthusiasm. Dixi Strand, Sisse Siggaard Jensen, Christian Jacquemin and Henning Christiansen also contributed greatly to the project.

REFERENCES

[1] Andreasen, T., H.Christiansen, J.Gallagher, C. Jacquemin, N.Møbius, N.Padfield, S.Siggaard, D.L.Strand, and P.R. Sørensen "Havhingstens Tur til Irland: en interaktiv oplevelsesplatform." CBIT, Roskilde University, 2011.

[2] Danks, M., M. Goodchild, K. Rodriguez-Echavarria, D.B. Arnold, and R. Griffiths. "Interactive storytelling and gaming environments for museums: The interactive storytelling exhibition project." *Technologies for E-Learning and Digital Entertainment, Second International Conference, Edutainment 2007.* Lecture Notes in Computer Science, 2007. 104-115.

[3] Falk, J.H. and Dierking, L.D. *Learning from museums: Visitor experiences and the making of meaning.* Alta Mira, 2000.

[4] Grigorovici, D. "Persuasive effects of presence in immersive virtual environments." *Being There: Concepts, effects and measurement of user presence in synthetic environments.* Ios Press, 2003.

[5] Jacquemin, C. "Architecture and experiments in networked 3d audio/graphic rendering with virtual choreographer." *Proceedings, Sound and Music Computing (SMC'04),* 2004.

[6] Juarez, A., W. Schonenberg, and C. Bartneck. Implementing a low-cost CAVE system using the CryEngine2. Entertainment Computing 1, 157–164, 2011.

[7] Krueger, M.W. "Environmental technology: Making the real world virtual." *Comm. ACM*, 1993: 36-37.

[8] MAX/MSP. 2011. http://cycling74.com/products/maxmspjitter/.

[9] Montemayor, J., A. Druin, A. Farber, S Simms, W. Churaman, and A. D'Amour. "Physical programming: designing tools for children to create physical interactive environments." *CHI.* 2002.

[10] OpenFrameworks. 2011. http://www.openframeworks.cc.

[11] Pinhanez, C.S., J. W. Davis, S. S. Intille, M.P. Johnson, A. D. Wilson, A. F. Bobick, and B. Blumberg "Physically interactive story environments." *IBM Systems Journal* 39, no. 3&4 (2000): 438 - .

[12] Sales Dias, J.M. "Natural and tangible human-computer interfaces for augmented environments." *Proceedings of the 26th annual ACM international conference on Design of communication, SIGDOC '08.* ACM, 2008. 181-182.

[13] Slater, M. and S. Wilbur. "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments." *Presence* 6, no. 6 (1997): 603-616.

[14] Sparacino, F. "Natural interaction in intelligent spaces: Designing for architecture and entertainment." *Multimedia Tools and Applications* 38, no. 3 (2008): 307-335.

 ISBN: 978-1-61208-170-0