SIMILAR Dreams

SIMILAR Dreams

SIMILAR Dreams

A European vision of multimodal interfaces in our future life

By the SIMILAR Network of Excellence Consortium

Edited by Sandra Daussogne and Benoit Macq

© Presses universitaires de Louvain, 2005

Registration of copyright: D/2005/9964/10

ISBN : 2-930344-91-1

Printed in Belgium

All rights reserved. No part of this publication may be reproduced, adapted or translated, in any form or by any means, in any country, without the prior permission of Presses universitaires de Louvain.

Distribution : www.i6doc.com, on-line university publishers
Available on order from bookshops or at
CIACO University Distributors
Grand-Place, 7
1348 Louvain-la-Neuve, Belgium
Tel. +32 10 47 33 78
Fax +32 10 45 73 50
Email duc@ciaco.com

” Everything you can imagine is real”

Picasso

Acknowledgement

SIMILAR is a European Network of Excellence in multimodal interfaces research.

It is sponsored by the EU in the 6th Framework Programme of the IST and is also known as FP6-507609 at the European Commission.

This book is the result of a close collaboration, mostly inside the consortium but also with outside partners, under the coordination of the editors.

Table of contents

ACKNOWLEDGEMENT **7**

TABLE OF CONTENTS **9**

CHAPTER 1: INTRODUCTION **19**

1.1. SIMILAR 19
1.2. COMPUTER ASSISTED HUMAN ACTIVITIES 20
1.2.1. Image-guided surgery *22*
1.2.2. Large Information Space visualization *23*
1.2.3. Trans-modality for handicapped persons and car drivers *24*
1.2.4. Edutainment *25*
1.3. SIMILAR PLATFORMS AND METHODOLOGY 25

CHAPTER 2: FUSION AND FISSION **29**

2.1. AUTHORS 29
2.2. INTRODUCTION 30
2.3. FUSION AND FISSION AT THE SIGNAL LEVEL 31
2.3.1. SP for Human multimodal signals fusion-fission *31*
2.3.2. Fusion-fission in medical applications *33*
2.3.3. Towards a Theoretical Framework *35*
2.3.4. Multimodal communications and distributed coding *36*
2.4. FUSION AND FISSION FOR DESIGNING HCI 36
2.4.1. Design of multimodal interaction and Fusion-Fission *37*
2.4.2. Context aware fusion and fission *40*
2.4.3. Fusion and fission in Natural Interaction Systems *41*

CHAPTER 3: OPENINTERFACE **43**

3.1. INTRODUCTION 43

3.2. STATE OF THE ART: SOFTWARE PLATFORMS 44
3.2.1. Multimodal data fusion 44
3.2.2. Software Engineering: Component-based platforms 45
3.2.3. Existing platforms/framework 46
3.2.4. Conclusions: Component-based platforms..... 49
3.3. GOALS 50
3.4. FEATURES 51
3.5. CONCEPTUAL ARCHITECTURE PROPOSAL 51
3.6. HOW IT WORKS..... 53
3.7. EDITOR 53
3.8. FUSION AND FISSION MECHANISMS..... 55
3.9. USERS..... 58
3.10. POTENTIAL APPLICATIONS 58
3.11. CONCLUSIONS 59
3.12. REFERENCES 59
3.12.1. State-of-the-art 59
3.12.2. OpenInterface..... 61

CHAPTER 4: MEDICAL APPLICATIONS 63

4.1. INTRODUCTION 63
4.2. MEDICAL STUDIO FRAMEWORK 64
4.2.1. Platform Architecture..... 64
4.2.2. GENERAL PURPOSE COMPONENTS..... 67
4.2.3. PROCEDURE-SPECIFIC SCENARIOS..... 69
4.3. OMICS DATA SUPPORT..... 73
4.4. REFERENCES 75

CHAPTER 5: BUILDING USABLE MULTIMODAL NIS 77

5.1. INTRODUCTION 77
5.2. STATE OF THE ART IN BRIEF 79
5.3. THE SIMILAR USABILITY SIG APPROACH..... 83
5.3.1. General objectives in brief 83
5.3.2. A bottom-up approach..... 85
5.4. APPLICATION DESCRIPTION STRUCTURE..... 86
5.4.1. Introduction..... 87
5.4.2. Application description 87
5.4.3. Technical issues..... 91

5.4.4. Functionality *94*
5.4.5. Interface and usability *97*
5.4.6. Evaluation *100*
5.4.7. Conclusions *102*
5.4.8. References *103*
5.5. NEXT STEPS 103
5.6. REFERENCES 105

CHAPTER 6: MULTIMODAL INTERFACES FOR PEOPLE WITH DISABILITIES 107

6.1. INTRODUCTION 107
6.2. MODALITY REPLACEMENT IN MULTIMODAL INTERFACES 110
6.2.1. Multimodal Interfaces *110*
6.2.2. Modality Replacement *110*
6.3. MAIN RESEARCH AREAS WITHIN SIMILAR 112
6.3.1. Haptic Applications for the Blind *113*
6.3.2. Haptic Tool to Access Presentations *113*
6.3.3. Virtual Interactive Training Environments for the Blind *117*
6.3.4. Sign Language Analysis *120*
6.3.5. Sign Language Synthesis *123*
6.3.6. Cued Speech Recognition *125*
6.4. THE DREAM SCENARIO 126
6.5. CONCLUSIONS 128
6.6. REFERENCES 129

CHAPTER 7: MIGRATORY MULTIMODAL UI 133

7.1. PROCESS 134
7.2. ARCHITECTURE 136
7.3. HIIS PROFILE 138

CHAPTER 8: SIMILAR WORK '04 143

1.1 INTEGRATION THROUGH ENTERFACE 144
1.2 INTEGRATION THROUGH EXCHANGES AND TWINNINGS 146
1.3 PUBLIC AWARENESS THROUGH SIMILAR WEBSITE 148
1.4 PUBLIC AWARENESS THROUGH THE SIMILAR NEWSLETTER 149
1.5 PUBLIC AWARENESS THROUGH CONFERENCES 150
1.6 PUBLIC AWARENESS THROUGH BOOKS AND PAPERS 153

1.7 PUBLIC AWARENESS THROUGH BEST SUMMER SCHOOL 154
1.8 PUBLIC AWARENESS THROUGH A SOFTWARE CONTEST 155
1.9 SIG ON INFORMATION FUSION AND FISSION 157
Multimodal sources representation in a HCI context 157
1.10 SIG ON USABILITY 159
1.11 SIG ON CONTEXT AWARE ADAPTATION 160
1.12 SIG ON MULTIMODAL ANALYSIS AND SYNTHESIS 162
1.13 SIG ON MEDICAL APPLICATIONS 163
1.14 SIG ON DISABILITY APPLICATIONS 164
1.15 SIG ON EDUTAINMENT APPLICATIONS 165
Virtual sculpturing 166
Artistic performance using 3D Body Tracking 167
Pointing interaction 168
Multimodal Speaker Identification 170
Surveillance and assistance in learning environments 170
1.16 COMMUNICATION BETWEEN PARTNERS 171
1.17 SUBMISSION OF EUROPEAN RESEARCH PROJECTS 173

CHAPTER 9: INTRODUCTION 177

9.1. THE "10-10-10" 177
9.2. INTERVIEWS 178

CHAPTER 10: TEN MAJOR ACHIEVEMENTS 179

10.1. RELIABILITY-ORDERED CASCADE 179
10.2. MUTUAL INFORMATION IN MULTIMODAL SP 180
10.2.1. Mutual information 180
10.2.2. Mutual information in multimodal signal processing ... 181
10.3. MULTIMODAL EMOTION DETECTION 182
10.3.1. Introduction 182
10.3.2. Achievements 183
10.3.3. Future research work 184
10.4. MULTIMODAL BIOMETRICS 184
10.4.1. Introduction 184
10.4.2. Different Fusion Strategies 185
10.5. MULTIMODAL MEDICAL IMAGE REGISTRATION 188
10.5.1. Image registration 188
10.5.2. Multimodal medical image registration 189

10.6. ICARE 190
10.6.1. Introduction: The CARE properties 190
10.6.2. ICARE: What is it? 192
10.6.3. Elementary Components 194
10.6.4. Composition Components 195
10.6.5. ICARE Platform 197
10.7. CAMELEON 199
10.8. MODALITY THEORY 202
10.8.1. MULTIMODALITY 206
10.9. USIXML 208
10.10. AMODEUS MODEL FOR DIALOGUE CONTROLLER 211
10.11. REFERENCES 213
Reliability ordered cascade 213
Multimodal emotion detection (prosody + facial) 214
Mutual information in multimodal SP 214
Multimodal biometrics 214
Multimodal medical image registration 214
ICARE 214
CAMELEON 215
Modality theory 216
UsiXML 216
AMODEUS 216

CHAPTER 11: 10 MAJOR PRACTICAL OUTCOMES 217

11.1. IMMERSIVE ENVIRONMENT 217
11.1.1. Introduction 217
11.1.2. The CAVE 218
11.1.3. Other immersive platforms 218
11.2. AIRPLANE COCKPIT 219
11.3. CAR APPLICATION 220
11.4. VTK-ITK MULTIMODAL MEDICAL IMAGING PLATFORMS 221
11.4.1. The Visualization ToolKit (VTK) 221
11.4.2. The Segmentation & Registration ToolKit (ITK) 222
11.4.3. Multimodal platforms 224
11.5. AUGMENTED REALITY IN SURGERY 226
11.5.1. Augmented reality 226
11.5.2. AR and surgery 226

11.6. MULTIMODAL MOBILE AUGMENTED REALITY 227
11.6.1. Definition 227
11.6.2. Applications 228
11.7. NICE DEMO 229
11.7.1. The NICE Project 230
11.7.2. The Multimodal Scenario 230
11.8. FAME: CONTEXT-AWARE DISTRIBUTED USER INTERFACES 232
11.8.1. What is it? 232
11.8.2. How does it work? 232
11.8.3. What is it used for? 233
11.9. CONCURTASKTREES 234
11.9.1. What is it? 234
11.9.2. Applications 236
11.10. GRAFIXML 237
11.11. REFERENCES 238
Immersive environment 238
Airplane cockpit 238
Car applications 239
VTK-ITK platforms for 3D multimodal medical imaging 239
Augmented reality in surgical operating theatres 239
Multimodal mobile augmented reality 239
NICE demo 240
FAME 240
ConcurTaskTree 240
GrafiXML 240

CHAPTER 12: 10 MAJOR CHALLENGES 241

12.1. SOCIAL INTERACTIONS 241
12.2. SOFTWARE ARCHITECTURE REFERENCE MODEL 242
12.3. MIX OF ARCHITECTURAL STYLES 243
12.4. SOFTWARE COMPONENT DESCRIPTION 243
12.5. POLYMORPHIC USER INTERFACES 243
12.6. GENERICNESS OF MODEL, METHODS, AND TOOLS 243
12.7. SCALABILITY 244
12.8. REFERENCE GLOSSARY 244
12.9. GENERIC CONTACTLESS MULTIMODAL INTERFACES 246

12.10. SECURITY/PRIVACY IN CONTEXT-SENSITIVE MULTIMODAL APPLICATIONS 246
12.11. REFERENCES 247

CHAPTER 13: INTERVIEWING 249

13.1. QUESTIONNAIRE 249
13.2. INDUSTRY 250
13.2.1. Xavier Marichal (Alterface, Belgium) 250
13.2.2. Frederic Kaplan (Sony, France) 252
13.2.3. Stéphane Chatty and Stéphane Sire (Intuilab, France). 253
13.2.4. Robin Springer (Computer Talk, USA) 261
13.3. ACADEMICS 269
13.3.1. Human-Computer Interaction 269
13.3.2. Context-aware adaptation 273
13.3.3. Usability 275
13.3.4. Information fusion & fission 285
13.3.5. Edutainment 298
13.3.6. Medical 304

GLOSSARY 313

FLAGS. 313
TERMS AND DEFINITIONS 313
REFERENCES 346

Part 1 :
Theoretical foundations

Chapter 1: Introduction

Benoît Macq (UCL, Belgium)

1.1. SIMILAR

The main specificity of SIMILAR is probably that it constitutes “the” European Network of Excellence where the Signal Processing technologies meet the Human Computer Interaction architectures.

Human Computer Interaction is achieved through interfaces. The digits in the Computer have to interact with the human senses, vision, hearing, speech, gestures, touching and even emotions that can be expressed in the facial expression or speech prosody. The very restricted typing of the keyboard and the pointing with a mouse have been the main ways of communication from the Human to the Computer while the windows of visual information (texts, graphics and images) displayed on the computer screen have been the usual channel of communication from the Computer to the Human. This way of designing human-computer interfaces has transformed humans and adapted them to the world of the digits. To enter this world, the laptop has to find a table and the computer users to leave the real world to enter in the cyberspace, fixing a screen and typing the keyboard in a quite autistic attitude. Of course, social interactions arise in this world also, through chats, role-playing games, etc. Such mode of interaction can be however extremely straining for impaired users. Should we accept a definitive adaptation of the human

senses to such restricted interactions to get access to the cyberspace and its wonderful collective intelligence? The two hands of a human have been slowly adapted through evolution steps of the human, to grasp, touch, show, express gestures to other humans. The vision of the human is stereoscopic and connected to a complex motor system. In “the purple rose of the Cairo” movie Woody Allen is entering in the screen to live in the virtuality of the story. Our aim is the contrary. We would like to contribute to the unleashing of the computers and let them move, interact and be present in a more efficient way in our real life. Their messages will be there to augment our perception of the world and the interactions with other humans, while we want to establish with them rich and multimodal interactions.

Recent progresses in speech processing, in computer vision and in scene composition (co-registration of virtual objects, being 3-D images, sounds or synthesized speech) have paved the way for offering breakthroughs in that direction. In parallel to these efforts in Signal Processing, a better understanding of the processes that happen when a human interacts with a computer is developed by the so-called “Human-Computer-Interaction” scientific community. Task modelling, dialog control according to a model of cognition, ergonomic criteria derived from cognitive psychology, adaptation of the modality, fusion of the various kinds of inputs and fission to the adequate channels at the output are key questions which are better understood and tested through usability procedures. Even more ambitious researches are addressed there through the modelling of dialog transformation thanks to the use of multimodal interfaces.

1.2. Computer assisted human activities

Mainframes and Personal Computers have been transformed from computation machines to supporting nodes for the human

collective intelligence through the extraordinary connectivity provided by the Internet. Collective intelligence is the capability for a group to organize itself in order to decide upon its own future and control the means to attain it in complex contexts. The World Wide Web is providing knowledge at your fingertips. Modelling this collective intelligence and how humans are interacting and evolving through it is a very complex research domain.

On the other hand, common objects of our usual life, are integrating more and more small computers for efficiently managing the human to machine communication and also regulate its behaviour. This is the case for mobile phones, cars, washing machines and so forth. For such situations, the tasks to accomplish are quite tractable and the improvements obtained through the intermediation of computers are measurable, through usability assessment procedures.

We decided to restrict our researches for cases where the task to accomplish can be more or less well modelled. We decided to address three application domains in this respect:

- Medical applications, where two major challenges are
    - multimodal image guided surgery and radio-therapy,
    - efficient visualization of very large information space in the biological continuum (from patient anamnesis to DNA going through organs, cells, molecules and proteins).
- Disabled persons and car drivers, for which the hands have to be replaced by other modalities like speech, facial features or even brain to computer direct interfaces (BCI).

- Edutainment, and specially the case where learning scenarios are applied in immersive environments.

The two first scenarios require **instrumental** interfaces while the last scenario can be more efficiently based on **anthropomorphic** interfaces generating natural interactions between the human and the machine.

1.2.1. Image-guided surgery

The surgeon's gesture is driven by what he sees (vision) and by the mechanical feedback he gets along the operation path. Minimally invasive surgery requires accessing the treated site with long flexible instruments introduced through natural openings or small incisions. Safe surgery needs complementary information such as the blood vessels, the nerves or other key connection paths to be identified because the path towards the target is only visible from the surfaces that appear along the incisions.

Augmented reality for image-guided surgery is a therefore a major driver for safe and minimally invasive operations. There is a strong push towards the design of advanced multimodal interfaces in the operating room (OR), to present multiple imaging modalities simultaneously to the physician, some of them being pre-operative and others intra-operative to track the gesture on duty. In these environments, interactions are based on techniques like augmented reality visualization, tracking technology, and voice input, and allow the surgeon to select among different information set ups and perspectives. Interactions have to be integrated to keep the hands free for the

surgical task. Complete interaction integration requires the design of a cognitive adequate system, whose purpose is to provide the right information at the right time to the surgeon. Those systems select the critical information related to the task, and present it in an augmented reality scene.

In order to optimise the signals to be delivered and the different kind of interaction to be performed with the images a task model of the operation has to be achieved, including planning, treatment and assessment. Criteria like continuity in the gesture, focus of attention, continuity in the interactions with accompanying nurses, are key factors to be assessed in the usability procedures.

Applications in the fields of neurosurgery and maxillo-facial surgery are presented in [Olsheywsky] and [Trevisan].

1.2.2. Large Information Space visualization

Recent progresses in the medical sciences have shown the interest for tackling a global biological continuum that provides clues about underlying clinical patterns for diagnosis and treatment of diseases, from global health statistics of a population to the DNA of a diseased cells, going through anatomical and molecular images of organs. Inferring patterns is only possible if appropriate information systems are properly chosen and inserted by MDs to appropriate biological models. The process of deriving from a large information space a diagnosis, a planning for the treatment and the assessment of it represents a key challenge for multimodal interfaces designers [Kitney].

A key benefit of multimodal interaction is increased scale of information access through conveying of high-level structure and specific details of relevance. This includes combining graphical modalities with sound and using multiple complementary graphical modalities (one for conveying high-level structure of the information space and another one for specific details) exploiting one or multiple display surfaces (wall, screen, PDA, etc.). In addition to output multimodal interfaces, input multimodality can facilitate the exploration and navigation within the large information space: examples include two-handed interaction to manipulate DNA structures and zoom maps.

1.2.3. Trans-modality for handicapped persons and car drivers

Usual commands of a car driver are occupying the two hands on the wheel and the two foots on the pedals. However new driving assistants are there to help through GPS to find a location, to draw attention on the car status or on potential risks on the road. Focus of attention of the driver is also a concern in the most recent systems. For these situations, the driver is focused on the car driving and has to access to information without the use of his hands.

Multimodal interfaces combining speech, visual augmentation through the windshield and haptic sensors on the wheel seems to offer promising prospects. In this situation the usual keyboard and windows interaction has to be translated in other modalities, taking into account the limited possibilities of the car driver. The study of such situations paves also the way for handicapped persons. SIMILAR is studying that kind of cases, including long-term research on brain to computer interfaces.

1.2.4. Edutainment

Edutainment consists in immersive interfaces which help the learning by playing.

The new multimodal interfaces allow reconstructing a space in which the user can be immersed and see and control his gesture for activating activities for learning or for entertaining. The sense of presence [of "being there"] is activated by the interpretation of the users in reaction to visual or audio signals generated by computing devices, for triggering physical actions in front of displays, for interpreting audio cues. Experiences are designed in a rich-media way in these immersive spaces, offering features that are integrated, as never any previous media could have done.

This implies not only the design of mixed reality spaces but also dialog control systems suited for learning and for story telling through natural interactions.

1.3. SIMILAR platforms and methodology

The main goal of SIMILAR is to bring together the Signal Processing and the Human Computer Interactions scientific communities into a single workforce to develop new methods for the design of multimodal interfaces. Fusion and fission of multimodal information will be a key scientific challenge. The group aims also at developing a software framework, named OPENINTERFACE and usability test procedures.

To tackle these challenges, the SIMILAR network of excellence articulated its works in eight mains Special Interest Groups.

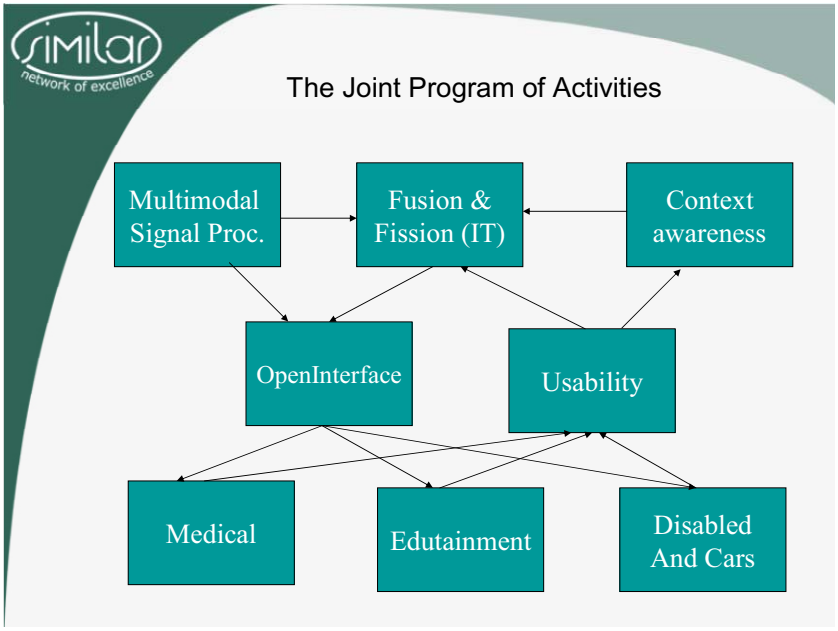


Fig. 1: SIMILAR scheme

The first SIG is devoted to new signal processing tools for multimodal interfaces. In this group the various signals corresponding to the human senses are explored. A strong involvement of teams in speech processing will have a major impact in natural interactions with computers through language.

A second major axis is the natural interactions through gestures. Recent progresses in computer vision allow an efficient pointing by a finger to a virtual object, grasping it, or more generally the natural search by gesture in very large information spaces. This implies a strong activity in gesture recognition but also the

creation of augmented reality interaction spaces, where the real scene is enhanced by synthetic objects or annotations. The speech prosody and the facial features analysis pave the way towards multimodal emotions recognition. The scientific society behind this activity is the EURASIP.

The second SIG group is related to context aware adaptation in the line of the works achieved in the CAMELEON project. Users expect to be able to employ their knowledge of a given version of the system when using the same service on another platform. Thus transitions between system versions and modalities of interaction have to be as smooth as possible. Many techniques and tools are required to develop user interfaces for multiple platforms.

The approach pursued in SIMLAR is the model-based paradigm. Starting from descriptions modelling tasks, users, and platforms, the tools can produce a set of platform-specific UIs. The added value of SIMILAR is to propose practical implementations and validations examples of the transformation rules. A specific effort is also addressed to provide inputs in standardisation bodies like W3C. Two examples of such ubiquitous applications can be found in the car driver case (the user prepares his trip on a laptop, goes outside with his PDA and then drives his car with a strong focus of attention on the way, going out his car with his PDA to finally find the place to go) or in the image-guided surgery (the surgeon prepares his intervention plans on pre-operative scanners, then he goes in the Operation Room where new images are acquired during the intervention and a mixed reality scene is recomposed from the high definition pre-operative plans to guide in optimal way the intervention).

A central SIG is the fusion-fission working group where a global approach of the management of modalities at the input of the

interface and at the output is developed. Fission is used at the input to extract the adequate features from the signals coming from each modality. At the output fission splits the information to be conveyed to the user to the channel adapted to his context and cognitive load. Fusion is achieved at the input to combine modalities according to their properties: Complementarity, Assignment, Redundancy, Equivalence (CARE properties). Fusion at the output is a key technology for the composition of mixed reality scenes.

The three applications domains are each covered by a dedicated SIG: medical, edutainment and disabled/cars.

The SIG devoted to usability aims at identifying well-defined, standard, frequently used, and "known-to-be-useful" evaluation methods, criteria and tools. These will form part of a first framework. New evaluation methods, criteria, and tools will be developed to assess the increasingly sophisticated and advanced systems SIMILAR wants to build. These methods/criteria/tools will progressively enhance the evaluation framework.

Low-level assessments, including precisions in pointing, false alarms vs. non-detection curves, error rates will be used for the assessment, particularly when modalities are combined to activate a command. This is particularly important for instrumental interfaces. Higher-level evaluation methods have to assess the cognitive load in the interaction but also the subjective aesthetic preferences of instrumental and anthropomorphic interfaces.

Chapter 2: Fusion and fission

2.1. Authors

This chapter, as chapters 10 to 12, is the result of the fruitful collaboration of the lab heads attending the Lausanne fusion – fission meeting held on 14th and 15th April 2005..

Its authors are thus (in alphabetical order):

- Michel Barlaud (I3S, France),
- Niels Ole Bernsen (NISLab, Denmark),
- Joelle Coutaz (UJF, France),
- Jean-Luc Dugelay (EURECOM, France),
- Benoît Macq (UCL, Belgium),
- Marcos Martín (U. Valladolid, Spain),
- Laurence Nigay (UJF, France),
- Fabio Paterno (ISTI, Italy),
- Michael Strinzis (ITI-CERTH, Greece),
- Murat Tekalp (Koç, Turkey),
- Jean-Philippe Thiran (EPFL, Switzerland),
- Dimitrios Tzovaras (ITI-CERTH, Greece),
- Jean Vanderdonckt, (UCL, Belgium),
- Slava Voloshynovskiy (U. Geneva, Switzerland),

2.2. Introduction

In the design and implementation of multimodal interfaces fusion and fission are playing a key role. Fusion is the process by which several channels of information, being low-level (signals) or higher-level (semantic information) are combined into a unique information entity. Fission refers to the decomposition phenomenon. Both fission and fusion processes have to be managed at the input and at the output of multimodal interfaces.

Considering fusion for the interpretation function, information channels often originate from distinct digital input channels or from distinct contexts. For example, several researchers of SIMILAR are working for emotion recognition in order to adapt the application to the emotional status of the user. The fusion process for emotion recognition can be done on low-level signals (facial expression, voice prosody, physiological signals as heart beats or brain activity) or at higher-level (based on linguistics, presuppositions of the user and behavior analysis).

Fusion is also a key issue for composing mixed reality scenes from several output channels. A key example for SIMILAR is the image-guided surgery case where pre-operative information and biological data can be used to enhance the real scene viewed by the surgeon. Co-registration of multimodal images and data are key-research topics within SIMILAR.

As for fission, it may be the case that information coming from a single input channel or from a single context need to be decomposed in order to be understood at a higher level of abstraction. At low level the fission process concerns various types of features extraction from a unique signal. An example in speech is the extraction of pitch, formants and energy. An example at higher level (from J. Coutaz) is the utterance “show me the red circle in a new window”. This sentence is received through a single digital channel but references two domains of discourse: that of the graphics task (i.e., “the red circle”) and that of the user interface (i.e., “a new window”). In order to satisfy the request, the system has to decompose the sentence into two

high level functions: “create a window” and “draw a red circle” in the newly created window.

Fission is a key-process in plastic user interfaces. An example is the guidance system for a car driver. The travel that was prepared on a classical laptop is fed in the car and the guidance splits the information into visual and speech channels and could provide even haptic stimuli in case of emergency. Such fission process is a key element for plastic user interfaces that adapt themselves to context changes.

2.3. Fusion and fission at the signal level

2.3.1. SP for Human multimodal signals fusion-fission

In many applications such as biometry, emotion recognition or command activation the input signals are the natural signals generated by a human, namely the speech, the gesture, the postures, and the facial expressions. A key role of signal processing for multimodal interfaces is to provide a clear feature extraction to get the most relevant information. Some exemplary features to be extracted in the case of human signals are

- For lip motion:
    - motion field
    - active contours
- For speech:
    - MFCCs (mel frequency cepstral coefficients)
- For voice
    - Formants (vocal resonance frequencies)
    - Pitch parameters (for prosody)
    - Energy (for volume)

- For gestures
    - Head,
    - Shoulder
    - Hand (rigid or non-rigid) motion parameters
- 3-D body attitude parameters from multiple views

The fission here is related to the extraction of relevant features from the signal. The features are generally used to feed a recognition model. The fusion is application driven.

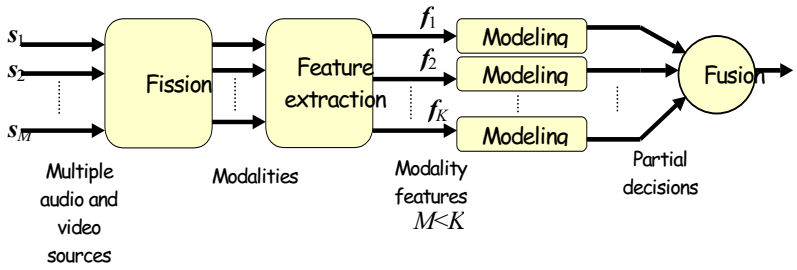


Fig. 2: Multimodal signal processing

Two among the main applications envisaged by SIMILAR are biometrics (user authentication: is it the right person in front of the interface?) and commands (commands of a car, of a cell phone or of wrist-watch computer application). The fusion (which can be cascaded in several steps) allows determining a specific command among a set of predefined ones or a specific user registered into a database.

The main measurement of performances of the process is based on reliability and confidence measure. Reliability is defined precise measure of the robustness against noise and acquisition artifacts. Confidence is related to a likelihood statistical analysis.

Receiver Operating Curves (RoC) of such systems measure the compromise that they are able to achieve between false alarms and false rejection of commands.

2.3.2. Fusion-fission in medical applications

Several Grand Challenges of medical applications require fusion-fission:

- **multimodal image segmentation**: segmentation is the process of extracting an object into an image by giving the same label to all the pixels belonging to the object. It can be done by classifying all the pixels with various labels (at least one label for the searched object, the so-called foreground, and one label of the remaining pixels, the so-called background) or by object delineation. In case of 2-D images the delineation can be achieved by an active contour and in the case of 3-D images the delineation can be performed by an active surface.

  SIMILAR is developing the case of segmentation based on multimodal images. A competition process which separates the object from the background is used to drive an active contour. The criterion used is the maximization of the mutual information of the pixels of the object across each modality. A complete mathematical framework is proposed by the team of Michel Barlaud in this respect on the basis of a variational approach called “Shape Gradients”. Segmentation can be viewed as a fission process, while the use of multimodal information to perform the segmentation implies a fusion process.

- **Multimodal image co-registration**: this is a fusion process which is important for diagnosis: e.g. a magnetic resonance image gives a nice anatomical view of an organ, while a positron emission tomography image will give a good view of abnormal activity (linked to the presence of a tumor for example) with poor resolution. The co-registration of such images is very important for diagnosis and treatment planning.

  A very important co-registration process is also related to image guided surgery where the co-registration allows creating mixed-reality images. The surgeon sees the scene in mixed-reality, which means that what he would see without the glasses is enhanced with synthetic information coming from pre-operative data. The mixed reality is very important in neurosurgery to avoid destruction of functional zones during a tumor resection operation. It can be used in maxillo-facial surgery to reach the targeted visage modifications. In co-registration the main concern is a fusion process, while in mixed reality the fission is important for the correct choice of the synthetic information added to the scene.

- **Tensors Signal Processing**: usual medical images are 2-D or 3-D images of intensities molecules. In Tensor imaging the values in the image are flow directions and intensities. At each pixel the flows of molecules is observed. In the brain, the connections between different areas can be viewed because the information is propagated by flows of water molecules. Tracking such flows allow determining fiber maps of the brain. The treatment of tensor imaging needs a fission process (features extraction from magnetic resonance images) and a fusion process (fiber tracking, in which the pixels belonging to a same pathway in the brain are linked together).

2.3.3. Towards a Theoretical Framework

The team of Jean-Philippe Thiran from EPFL attempted to set a framework for the feature extraction issue in multimodal application. The proposed framework concerns both the fusion of various modalities. The input modalities are observations of a common reality (what we call in Signal Processing a “scene”):

- **Example 1**: the speech signal and the video signal from a scene with several persons: the multimodal treatment allows to extract features from the video which allows to position the speaking person in the visual scene

- **Example 2**: the different modalities (RX scanner, Magnetic Resonance and Positron Emission Tomography) of a human organ (like the brain): features extraction allows to better co-register the different modalities into a single image containing the fused information of each modality.

The goal is to optimize the feature extraction process in order to minimize errors about the observation of the scene. It has been shown in this work that features of each modality that maximize the mutual information between modalities are the ones which minimize the probability error when the observation modalities can be linked to the real scene by a Markov chain model (i.e. an observed status is linked to the real scene with a simple probability transitions model).

2.3.4. Multimodal communications and distributed coding

Several researchers of SIMILAR are studying the case of multi-user communications over discrete memoryless networks. Key questions are how to compromise various modalities in a global rate distortion framework (fission is here the weight to provide to each element of a global transmitted scene to ensure graceful degradation when the rate drops). In distributed coding optimal channel and source decoding can be done by taking into account the redundancies existing between each transmitted element of the scene: the decoding uses these redundancies to augment the quality of each element in a kind of fusion process.

2.4. Fusion and fission for designing HCI

Two main cases of multimodal interfaces can be considered:

- **Instrumental interfaces** for which a specific task model exists and for which it can be envisaged to have precise evaluation procedures.

- **Anthropomorphic interfaces** for which the goal is not a precise achievement of a measurable task but more to provide a natural interaction. This approach is mainly developed at NISLab.

2.4.1. Design of multimodal interaction and Fusion-Fission

The global design space proposed by Laurence Nigay is depicted here below

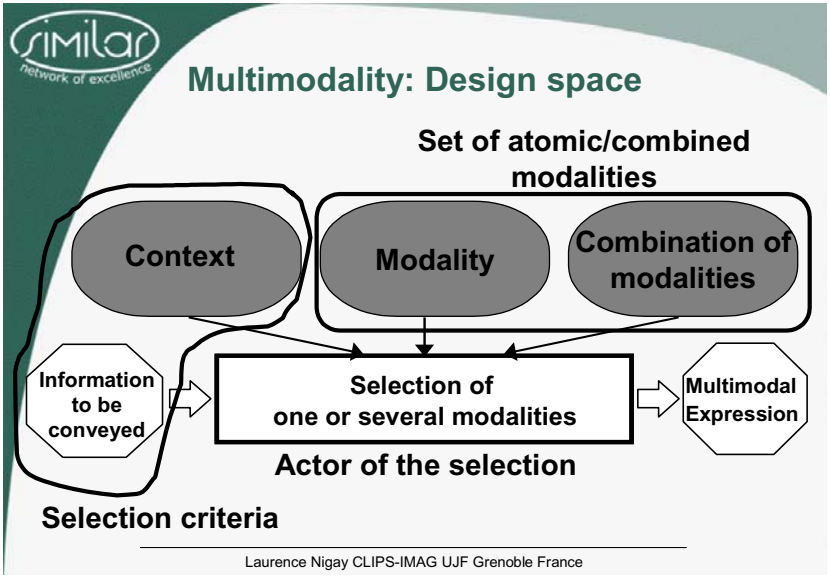


Fig. 3: Multimodality design space

The inputs of the system are the information to be conveyed which can be low-level or at the semantic level. It could be man-machine information (commands), man-media (natural interaction for edutainment applications) or man-man information (telepresence). The optimal choice of a modality depends on the context. Using the speech modality for communicating during a concert is clearly not a good choice, while it could be the best choice for a surgeon in the operation room (OR).

A key point is the selection of the modalities both at input and output of the interface:

- It can be fixed before the system is running;
- It can be done at run-time by the system itself;
- It can be done at run-time by the user.

The optimal (combination of) modality for a particular activity can be very user-dependant and can vary in time (after a learning period). Assessment can be done by a so-called wizard of oz where the system adaptation (choice of modality) is done by a human observer in replacement of the system, but giving the sensation to the user that the adaptation is only done by the autonomous computerized system itself.

The most important factor influencing the choice of a particular modality is the context. Context awareness and adaptation is therefore a key success factor in designing multimodal interfaces.

A modality is defined by a couple (device, interaction language). Some modalities are active (activated explicitly by the user, e.g. mouse click), some others are passive (activated implicitly by the user, e.g. head tracking).

One of the most interesting theory related to language interaction of modalities is the modality theory of N. O. Bernsen.

Several models are handling the properties of combinations of modalities (TYCOON, CARE, …). The CARE properties are defined as relationships between Devices, Interaction languages and Tasks. The combination of modalities can be as follows:

- C : Complementarity
- A : Assignment
- R : Redundancy
- E : Equivalence

In the case of Assignment the modality is directly connected to a particular command. In the case of complementarity, redundancy and equivalence a fusion process is required. It can be done at a low level on the signal itself or later at a higher semantic level (late fusion).

An important step in describing multimodal interfaces from a system point-of-view, is a conceptual separation of the devices, the language and the associated tasks.

A vision proposed by Laurence Nigay is as follows: the fusion can be done at different levels of abstraction.

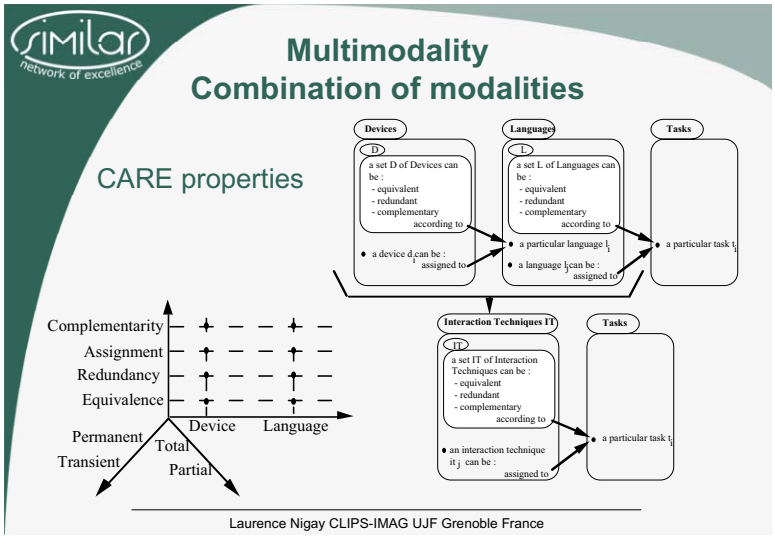


Fig. 4: Levels of abstraction

In ICS (Interacting Cognitive Subsystems), the human information processing system is subdivided into a set of specialized subsystems. The sensory subsystems transform sense data into specific mental codes that represent the struc-

ture and content of incoming data. These representations are then handled by subsystems specialized in the processing of higher-level representations: the morphonolexical subsystem for processing the surface structure of language, the object subsystem for processing visio-spatial structures, and the prepositional and implicational subsystems for more abstract and conceptual representations. The output of these higher-level subsystems are directed to the actuator subsystems (articulatory and limb).

ICS can be used for predicting cognitive resources involved in using and choosing modalities. A dialog controller AMODEUS has been developed on this basis in frame of an ESPRIT European project (www.mrc-cbu.cam.ac.uk/amodeus/).

Nowadays the ICARE and OPENINTERFACE projects aim at developing a toolbox for an easy component-based development of interfaces. The fusion of modalities is done at the design phase.

2.4.2. Context aware fusion and fission

The dynamic adaptation of multimodal interfaces has to be done according to the context of use: platform, physical environment and user. It is important to clearly define the appropriate levels of abstraction to reason about fusion and fission. For this purpose Joëlle Coutaz proposes the following layered representation for context acquisition in multimodal interfaces application:

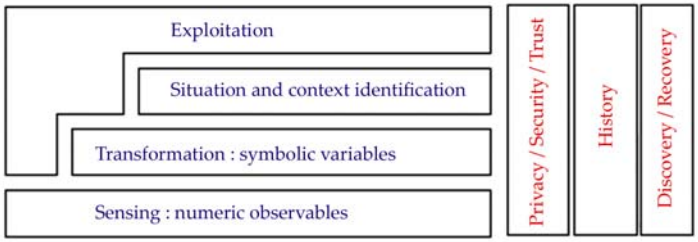


Fig.5: Layers of context acquisition

Research in SIMILAR aims to built fusion & fission techniques that are self-reconfigurable and for which the adaptation to context should be measurable in terms of

- Robustness and uncertainty of data (from sensors in particular, arrival/departure of resources)
- Ambiguity
- Latency (from the user's perspective)
- Awareness of the user (including security/privacy concerns)

The developed fusion/fission mechanisms should be built for each level of abstraction, then look at the interplay between multiple levels of abstraction. The interoperability and the temporal coherency of the systems are key factors to keep in mind.

2.4.3. Fusion and fission in Natural Interaction Systems

Niels Ole Bernsen presented a system, the NICE system, which aims at developing an anthropomorphic interactive system that is demonstrated for an application where the user can discuss in a

natural way with Hans Christian Andersen. The proposed architecture of the NICE project is as in the next figure:

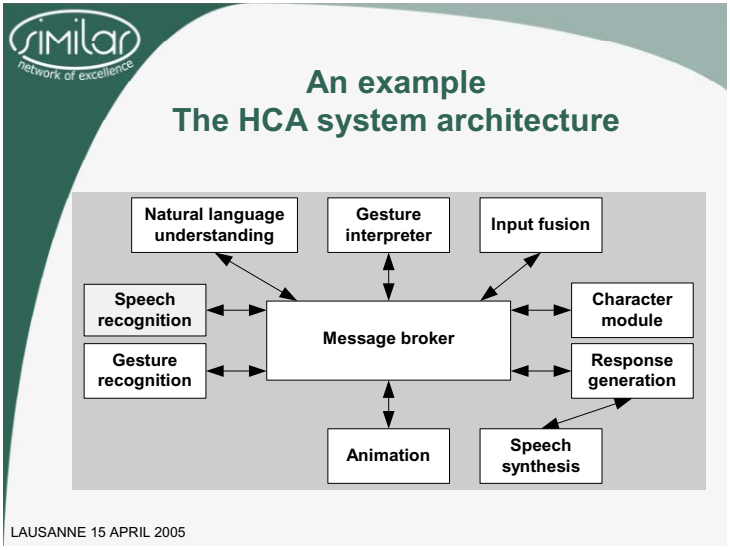


Fig.6: NICE architecture

The works of NICE stressed some important challenges to be addressed by SIMILAR:

- Semantic fusion of language, vision, touch, etc. input
  - we have only just started on this one
- Ontology-based learning, reasoning, planning in environmental and communication contexts
  - A large amount of work is still needed here in the next years.
- Ontology-based automated output generation
  - from concepts and goals, not just from templates

For NISLab the long-term goal of multimodal interfaces is to contribute to the building of a virtual human!

Chapter 3: OpenInterface

Lionel Lawson
& Daniela Trevisan (UCL, Belgium)

3.1. Introduction

OpenInterface is the SIMILAR software platform that includes software components dedicated to multimodal interaction and multimodal data fusion. OpenInterface is a software platform integrating results from the Human-Computer Interaction (HCI) community as well as from the Signal Processing community.

Indeed the HCI community contributes with:

- interaction modalities (for instance, direct manipulation or orientation/location tracker),
- software mechanisms for multimodal interaction and in particular fusion of modalities.

The Signal Processing community provides:

- algorithms for multimodal data fusion (for instance, a segmentation algorithm),
- algorithms and fusion mechanisms as being building blocks for defining interaction modalities (for instance, a vision tracker).

The approach is not to impose a structure for an API (Application Programmer Interface) or to restrict the hardware and software base, for this approach has already been tried and doesn't work. OpenInterface is far more open than that. It will cover all possible interactions between ranges of devices from large computers to portable phones using standards whenever possible, such as those provided by W3C[1].

3.2. State of the art: Software platforms

Currently there is no widely supported industry standard for multimodal applications, which is the key to widespread adoption.

3.2.1. Multimodal data fusion

Nowadays there are several toolkits, libraries and frameworks supporting the multimodal data fusion. Toolkits and libraries are frequently used for signal processing in general and frameworks architectures for integration of components and development of final applications.

Libraries such as MatLab [1], VTK [2], ITK [3], VXL [4], LTI-Lib [5], MITK [6] are potential candidates to be included as components into OpenInterface Platform once they are capable to provide the appropriate modalities treatment.

Some Medical frameworks such as Medical Studio [0], VAVFrame [8] and Sinergia [9] rely on the participation of SIMILAR partners and the potential merge with OpenInterface is envisioned in a second phase of OpenInterface development (with copyright licenses implemented).

[1] World Wide Web Consortium (www.w3.org) is currently working on the development of a multimodal standard.

Medical Studio [7] is a cross-platform component-based architecture supporting only C and C++ components and designed for programmers and end-users. The VAVFrame [8] is not cross-platform, nor does it support heterogeneous components. It is highly dependent on Microsoft resources. On the other hand the VAVFrame user's architecture corresponds to the same three user levels supported in OpenInterface (i.e. application designers, components programmers, and end-users).

3.2.2. Software Engineering: Component-based platforms

The OpenInterface work group focuses on component oriented platforms/frameworks, more specifically on the interoperability mechanism used for establishing communication among heterogeneous software components. The latter can be either existing components or new components that have to be integrated in a component pool.

There are numerous component based models/platforms/frameworks or what have you. This state of the art aims at technologies that have already been tested/implemented (UniFrame [10], CCA [11], Babel [12]) and/or at widely recognized technologies (such as CCM [13], EJB [14], COM+ [15] etc.). Three main integration technologies are CORBA, EJB, and COM.

- The CORBA Component Model (CCM) [13] is a specification of a server-side component model for building and deploying enterprise-class applications.

- EJB [14] is intended to support distributed Java-based enterprise-level applications, such as business information management systems. Among other things, it prescribes an architecture that defines a standard, vendor neutral interface to information services including

transactions, persistence, and security. It thereby permits application developers to create component-based implementations of business processing software that are portable across different implementations of those underlying services.

- The Component Object Model (COM) [15] is Microsoft's object model and middleware platform for distributed object computing and component-based development. As an object-oriented middleware platform, it is similar to CORBA in concept and operation.

These techniques/technologies are widely used, mainly for the construction of development environments aimed at the easy integration/re-use of heterogeneous components. With respect to the development of the low-level integration part in OpenInterface, they all have drawbacks.

More generally, the construction of a development environment requires deep knowledge of such technologies. They are all also intrusive, Object-only integration techniques. That means that the software parts necessary for the integration of components are tightly bound to the components' implementation. This therefore demands re-engineering and a substantial amount of code modification.

3.2.3. Existing platforms/framework

Here we are focussing on component-oriented development frameworks/platforms aiming at the reusability and inter-operability of software components. Most of the materials presented here are still ongoing research projects.

3.2.3.1. Common Component Architecture Forum

The objective of the CCA Forum [11] is to define a minimal set of standard interfaces that a high-performance component framework has to provide to components, and can expect from them, in order to allow disparate components to be composed together to build a running application. Such a standard will promote interoperability between components developed by different teams across different institutions.

First of all, at a low-level interactions point of view, CCA addresses the same issues as OpenInterface. It also solves them using the same principles. But at a higher level, CCA lacks the signal processing, multimodalities, and context-aware component properties that OpenInterface intends to embed.

Then, it is worth noting that current CCA framework implementation is not very mature. It is still under development or in the improvement phase and is subject to evolution and major changes. Therefore developing a new application based on that tool can be a tough task because of software immaturity.[0]

3.2.3.2. UniFrame

The goal of UniFrame [10] is to help build interoperable distributed computing systems. UniFrame framework aims at seamless interoperation of heterogeneous distributed software components. By using UniFrame, a new system can be built by assembling pre-developed heterogeneous and distributed software components.

The goal of UniFrame is to facilitate interoperability between components built on top of different component models. The framework does not address performance issues when dealing with communication between different programming languages;

this is why SOAP is used as a common language for communication among components.

Finally, UniFrame does not include a generic way of adding new properties to components. For instance, the first versions of OpenInterface should focus on Context-aware and multimodalities components. Those components need specific attributes to be considered as such (i.e. context-aware or multimodality).

3.2.3.3. Schooner

The goal of the Schooner Project [17] is to investigate a new model for High Performance Computing (HPC) applications in which the application is constructed as a heterogeneous distributed programme or, equivalently, a meta-computation. With Schooner, an application is constructed from a distributed collection of component codes executing on a variety of heterogeneous machines. A software interconnection system connects components into a single programme in a manner that is transparent to the user, and provides configuration and execution control over the resulting computation.

Schooner shares several goals with OpenInterface, such as the construction of software from heterogeneous component assembly and runtime computing efficiency.

However, there are some limitations:

- First, integration of existing components in Schooner requires substantial modification of code. OpenInterface aims to minimise the amount of code modification needed to integrate third part components.
- Secondly, Schooner control flow is based on a sequential procedural programming model. In OpenInterface, parallel computing is a main issue
- Thirdly, Schooner does not have a way to add domain-specific properties to components.

3.2.3.4. ArchStudio

ArchStudio [18] is an architecture-driven software development environment. It focuses on software development from the perspective of software architecture. ArchStudio designers use Software architecture as a way of looking at systems made from independent components and connectors that are linked together in one or more configurations.

Briefly, ArchStudio does not address the heterogeneous languages interoperability. It focuses on architecture-directed application design. The latter approach would be useful to re-use at a later stage of OpenInterface development, i.e., high-level application production.

3.2.4. Conclusions: Component-based platforms

The tools and technologies presented above are used for component-oriented programming and more specifically for heterogeneous component interactions.

Two categories of tools were presented, namely, generic component-oriented standards (such as CCM, EJB, and COM) aimed at seamless interaction, and existing component-oriented platforms.

As a result of the state of the art on component-based platforms it transpires that at lower point of view (efficient heterogeneous component interaction) none of the tools (i.e. IDL, proxies, stubs, factories, etc.) suits the OpenInterface needs, even though their underlying principle are the same. On the other hand the libraries and tools covering the signal processing area should be integrated as components into OpenInterface Platform. More details about OpenInterface Platform goals and requirements are described hereunder.

3.3. Goals

The main contributions of OpenInterface can be summarised as follows:

- The OpenInterface platform facilitates the development of multimodal applications. New tools and developed techniques are registered in the platform as re-usable, interconnectable components. An executable application scheme is easily created by using the components registered in the platform.

- Redundancies of common tasks while developing multimodal applications (e.g. data formats support, usual audio/image/video processing, etc.) are avoided.
- Easy integration of components into the platform using a Common Interface Description Language (CIDL). It should be compatible with USIXML (www.usixml.org) assuring in this way the context-aware adaptation process. USIXML consists of a User Interface Description Language (UIDL) allowing designers to apply multidirectional development of user interfaces at multiple levels on independence, and not only device independence. This descriptive language is being progressively adopted in SIMILAR.
- A common platform that will let partners focus on their work and not on tasks that have already been done by others (e.g. image reading/visualization, video capture, etc.)
- Rapid prototype and testing usability and ergonomic aspects during the development of multimodal applications.
- Connection point between two different communities, Signal Processing and Human-computer Interaction.

3.4. Features

- Component-based architecture
- Heterogeneous native components
- Easy integration of components
- Connection between components to develop new multimodal application
- Ergonomic properties of fusion and fission components (ICARE concepts)
- Cross-platform development
- Support for different copyright licenses

3.5. Conceptual architecture proposal

OpenInterface adopts the component-based architecture concepts and its functional decomposition follows a client-server architecture. Three functional units have been identified, namely, client, dispatcher, and server units. The functional units' interactions are described in Figure 7.

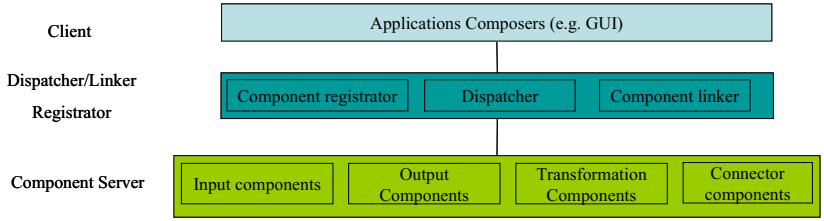


Figure 7. Functional unit interactions of OpenInterface platform.

Each functional unit is broken down into several components conforming to the following sections.

Component server

Component server corresponds to the platform repository. It is composed of four component classes, *i.e.* input, output, transformation and connector. All potential components to be integrated into the platform will fit in one of those classes. Examples of such components are described below.

- Input Components (source): microphone driver, mouse driver, voice analysis with bundled sound capture library, etc.
- Output Components (sink): display tools, voice synthesis, etc.
- Transformation Components (filter, transformation): bundled fusion & fission, translator, voice analysis without bundled sound capture library, miscellaneous algorithms, high-level communicator, etc.

Dispatcher/Linker/Registrator

This functional unit has three main services:

- Component Registrator: "yellow page"
- Component Linker: pipes, procedure call, etc…
- Dispatcher: allows one to create the configuration schema of the components composition

This functional unit aims at the automatic generation of *platform specific components* (i.e. OpenInterface components) from specifications (i.e. CIDL) and code.

Client

This functional unit provides composition facilities such as providing the appropriate translation tools for connecting components with incompatible output/input types.

3.6. How it works

Each component is registered in the OpenInterface Platform using the Component Interface Description Language (CIDL), which is described in XML. See the flowchart in Figure 8. The registered components' properties are retrieved by the Graphics Editor (Java). Using the editor the user can edit the components properties and compose the execution pipeline (by connecting the components) of the multimodal application. This execution pipeline is sent to the OpenInterface Kernel (C/C++) to run the application.

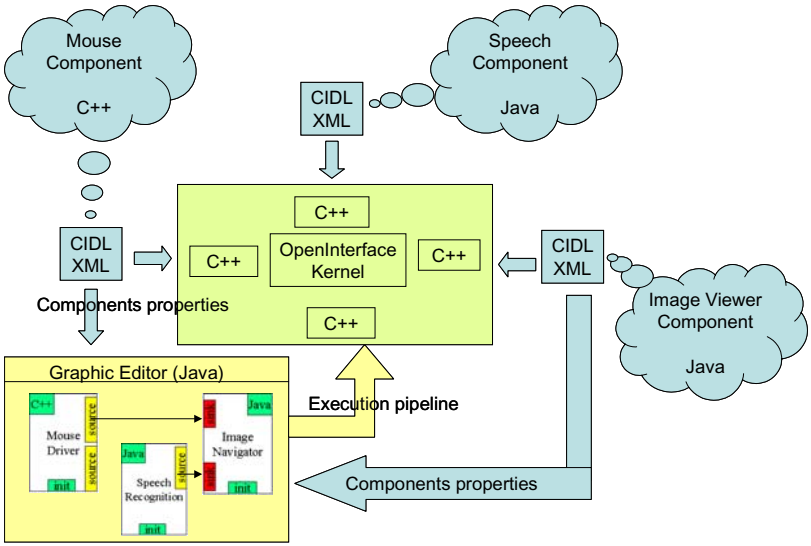


Figure 8. Overview of the OpenInterface Platform

3.7. Editor

The objective is to facilitate the creation of multimodal design with an easy and intuitive user interface called OpenInterface Editor. It will be useful for Application Designers (AD).

The application main functionalities are to:

- assemble components from Signal Processing and Multimodal Interaction.
- save the result in an XML format that can be used to compute it.
- open an existing project and make modifications.

The editor is composed of:

- 3 panels:

– a components panel that contains all the components available, according to the working mode (Signal Processing or Multimodal Interaction).

– a list panel that contains the list of the project components.

– a property panel that contains the list of properties for the selected component.

- 1 tabbed panel that contains the different working panels. When edited, each assembled box is opened in a new tabbed panel.

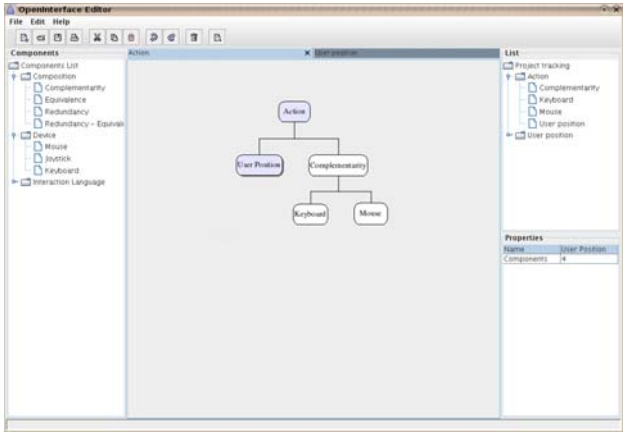


Figure 9. OpenInterface editor-graphic interface.

3.8. Fusion and Fission Mechanisms

The Framework should also include some mechanisms to handle how the multiple interaction modes (which are inserted as *plug-ins*) might work together in the future.

In multimodal systems, an event is a representation of some asynchronous occurrence of interest to the multimodal system, such as mouse clicks, hanging up the phone, speech recognition results, and errors. Events may be associated with information about the user interaction, such as where the mouse was clicked. Interactions (input, output) between the user and the application may often be conceptualized as a series of dialogues managed by an interaction manager.

So far, in designing multimodal applications the designer needs to specify the multimodal interaction dedicated to a given task of the interactive system under development. The concepts approached in the ICARE platform (Bouchet, J., Nigay, L., 2004) should handle these fusion (multimodal inputs) and fission (multimodal outputs) issues, which will be integrated into OpenInterface Framework.

ICARE stands for Interaction-CARE (Complementarity Assignment Redundancy Equivalence). Two kinds of ICARE component are considered: (1) elementary components that enable the designer to define “pure interaction modality” as defined in the theory of modalities (Bernsen, N., 1994), and (2) generic composition components that enable the designer to specify combined usage of modalities. Unlike elementary components, composition components are generic in the sense that they are not dependent on a particular modality.

The user of the ICARE platform selects the modalities and specifies the combination of modalities in terms of the CARE ergonomic properties (Nigay, L., Coutaz, J., 1997), all by

graphically assembling software components without knowing the details of the components' codes.

A first multimodal system developed with ICARE components is MEMO, a GeoNote system (Persson, P., Espinoza, F., Cacciatore, E., 2001). MEMO allows users to annotate physical locations with digital notes that have a physical location and are then read/removed by other mobile users. Figure 4 shows the MEMO ICARE specification for input multimodal interaction. Two tasks are possible using the modalities. They define what the rest of the system receives from the ICARE components: (1) orientation and localisation of the user (T1), so that the system is able to display in the HMD the visible notes according to the current position, and orientation of the mobile user (2) manipulation of a note (create, pick and remove a note) (T2).

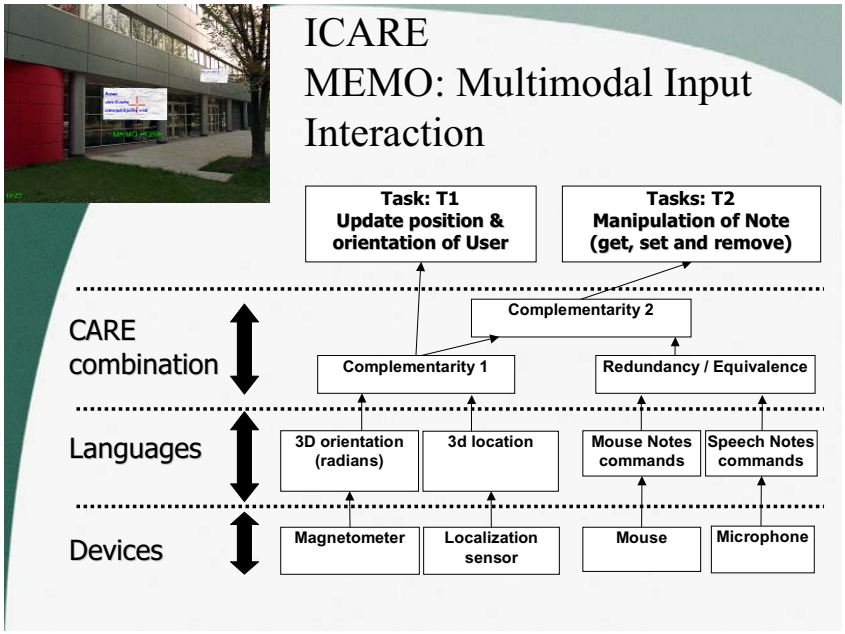


Figure 10. ICARE specification of MEMO input interaction.

One modality, “orientation”, is represented by the couple magnetometer (Device component) and the three orientation angles in radians (Language component), another modality, “localisation”, is represented by the couple (Localization sensor, 3D location). The modalities “orientation” and “localization” are complementary (Complementarity-1 component). Two equivalent modalities are dedicated to the manipulation of the notes, *i.e.*, commands specified using a mouse and speech commands.

Based on the ICARE specification of Figure 10, every three milliseconds, the modality “orientation” provides a vector of three floats corresponding to 3D orientations in radians (yaw, pitch, and roll). With the same frequency, the modality “localization” provides a vector of three floats corresponding to the user’s 3D position (x, y, z). The Complementary-1 component effects the fusion of these two vectors. An event is triggered s soon as the vector of six floats is complete (if an eager strategy is used) or when the temporal window is finished (if a lazy strategy is used), and the vector is passed to the next component, namely Complementary-2. If the timestamp of the event corresponding to the command <remove> that was received from one of the two equivalent modalities belongs to the same temporal window to which the six-float vector event belongs, the two events are combined. The component Complementary-2 then sends the complete command <remove, six parameters> to the system’s Dialogue Controller, which will determine the corresponding note to be removed based on the set of notes stored in the Functional Core. The complete code of the interactive system is structured along the ARCH software architectural model (UIMS Tool Developers Workshop, 1992).

With this approach, automatic checking of ergonomic properties can be supported while the designer is specifying the multimodal interaction. For example action continuity (Dubois, E., Nigay, L., Troccaz, J., 2002) can be checked automatically based on ICARE Device component properties.

3.9. Users

OpenInterface is designed to serve three levels of users, namely, programmers, application designers (AD), and end-users. Programmers are responsible for the development and integration of new components into the platform. Application designers are low-level users aware of end-user's needs and the resources provided by the platform. Basically the AD will make use of the graphics editor to compose the final multimodal application. End-users do not have access to the application code; they just interact with the final application interface.

3.10. Potential applications

Benchmarking

OpenInterface can be used to facilitate benchmarking algorithms and/or applications. As an example, a method to perform registration between two images could be done using different feature extractors and interpolator components. One interpolator component developed in C++ could be replaced by another developed in MatLab, thereby facilitating the comparison of results.

Medical, disabled and edutainment applications

Despite their differences, these applications have common visualisation and processing needs. On the other hand, for instance, an automatic method for registering medical images can be designed interactively thanks to the multimodal OpenInterface components. OpenInterface aims to avoid this redundancy and to speed up end-product development through the development of a flexible software environment

Prototyping

OpenInterface will allow rapid prototype and testing usability and ergonomic aspects during the development of multimodal

applications. For instance, a driver simulator able to react appropriately to hypovigilence could easily be developed connecting different OpenInterface components. Such components include acquiring physiological signals, analysing facial expression, and tracking the user's focus (eye tracking) in an augmented reality view.

3.11. Conclusions

We believe that the wide acceptance of the open-source OpenInterface will be a key factor of SIMILAR's success, reducing the current gap between human- computer interaction factors and the signal processing community. It will make the participation and contribution of external partners easy and attractive by narrowing the gaps between developers, application designers, and end-users inside and outside the network community.

3.12. References

3.12.1. State-of-the-art

Matlab Software. [Online]. Available at www.mathworks.com/

Schroeder, W., Martin, K. and Loresen, B. The visualisation toolkit. Kitware, 2002.

Schroeder, W. and. Ibanez, L . The itk software guide. The Insight Consortium.

VXL [Online]. Available at http://vxl.sourceforge.net/

LTI Lib [Online]. Available at http://ltilib.sourceforge.net/doc/homepage/index.shtml/

MITK – Medical Imaging Interaction Toolkit [online]. Available at http://www.mitk.org/

Medical Studio. [Online]. Available at www.medicalstudio.org/

VAVFrame [Online]. Avalable at http://www.vavlab.ee.boun.edu.tr/ProjectsNEW/links_VAVframe.html

Monserrat, C., López, O., Meier, U., Alcañiz, M., Juan, C., and Grau, V. GeRTiSS: A Generic Multimodel Surgery Simulation, Proc. Surgery Simulation and Soft Tissue Modeling Interntaional Symposium, Juan-Les-Pinnes (Fr), Lecture Notes in Computer Science 2673, pp. 59-66, 2003. (SINERGIA)

Zhao, W., Bryant, B. R., Raje, R. R., Auguston, M., Burt, C. C. and Olson, A. M. Automated Glue/Wrapper Code Generation in Integration of Distributed and Heterogeneous Software Components, Proc. of the 8th IEEE International Enterprise Distributed Object Computing Conference (EDOC2004), IEEE Press, pp. 275-285, Monterey, California, USA, 2004. (UNIFRAME)

CCA - The common component architecture forum. [Online] Available at www.cca-forum.org/.

Babel, a language interoperability tool for high-performance parallel scientific software. [Online] Available at www.llnl.gov/CASC/components/babel.html.

Corba component model. [Online] Available at www.componentsource.com/ Services/-CORBAWhitePaper.asp.

Anna Eriksson. Integration of heterogeneous systems, a component-based approach. Master's thesis, Department of Informatics. School of economics and commercial law at Gothenburg University., 2001. (EJB)

Pryce, N. G. Component Interaction in Distributed Systems. PhD thesis, Imperial College of Science, Technology and Medecine. University of London: Departement of Computing, 2000. (COM/COM+)

Diva/CCA coding camp. [Online] Available at www-vis.lbl.gov/ Events/DiVA-July2004/findings.html.

The schooner project: Interconnecting heterogeneous hpc scientific applications. [Online] www.cs.arizona.edu/schooner.

Archstudio, an architecture-driven software development environment. [online] Available at www.isr.uci.edu/projects/ archstudio/whatis.html.

3.12.2. OpenInterface

Bernsen, N. (1994). Modality Theory in support of multimodal interface design. Proc. of Intelligent Multi-Media Multi-Modal Systems, p. 37-44.

Bouchet, J., Nigay, L.(2004) ICARE: A Component-Based Approach for the Design and Development of Multimodal Interfaces, Extended Abstracts of CHI'04, Vienna, Austria, April 2004, pp 1325-1328.

Dubois, E., Nigay, L., Troccaz, J. (2002). Assessing Continuity and Compatibility in Augmented Reality Systems. UAIS Journal, 4, p.263-273.

Nigay, L., Coutaz, J. (1997). Multifeature Systems: The CARE Properties and Their Impact on Software Design. Intelligence and Multimodality in Multimedia Interfaces, AAAI Press.

Persson, P., Espinoza, F., Cacciatore, E. (2001) GeoNote: Social Enhancement of Physical Space. Proc. of CHI2001 p.43-45.

UIMS Tool Developers Workshop (1992), A Metamodel for the Runtime Architecture of an Interactive System. SIGCHI Bulletin p. 32-37.

Chapter 4: Medical applications

Monica Gemo (UCL, Belgium)
and Richard Kitney (Imperial College, UK)

4.1. Introduction

Advances in scanning technology and other data collection systems deliver a wide spectrum of useful and complementary information about a patient's status to research and clinical domains. Diagnostic information is present at a range of dimensions from system and organ down through cellular to the molecular structure of proteins and genes and finally the gene sequence. This range of scales is referred to as the Biological Continuum. New technologies, which will become increasingly available in the post-genomic era, will provide the clinician with more effective "tools" to assess the individual patient's state accurately.

To exploit these possibilities to their full potential, while avoiding overwhelming clinicians with a flood of information, especially in a critical environment such as the operating theatre, diagnostic and treatment functionalities need to be coupled to advanced, intuitive user interfaces. At the same time, these functionalities need to build upon common platforms and tools to permitting the

sharing and processing of the spectrum of imaging modalities and the scales of the Biological Continuum.

4.2. Medical Studio Framework

Medical Studio is a multimodal component-based platform that supports a complete medical processing-pipeline, including data processing, visualisation and navigation in pre-operative planning and intra-operative guidance.

Its modular architecture can easily manage abstractions of hardware peripherals and make data directly available from them. The multimodal feature refers to both image fusion processing and intuitive multimodal modes of interaction with the final application. Components developed in collaboration with several research centres and medical clinics have shown the promising dissemination and versatility of this medical framework in various disciplines. Such components provide support for transcranial magnetic stimulation, segmentation, image registration, and augmented reality visualisation, among other things.

4.2.1. Platform Architecture

A framework allowing the centralisation of all assisted surgery tasks must have a consistent and evolvable architecture. The software architecture implemented in Medical Studio is a modified version of the ARCH [1] model to add a specialisation of the input and output modalities. The principal components of the architecture illustrated in Figure 11 are detailed as follows:

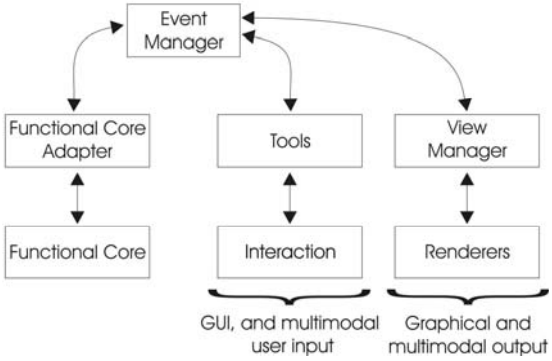


Figure 11 Medical Studio architecture.

- **Functional Core** contains all data processing components such as registration algorithms, I/O filters, etc.
- **Functional Core Adapter** is an abstraction layer that allows the event manager to communicate with the functional core.
- **Event Manager** is the dialogue controller between interactions and functional core. When interactions occur, the event manager will propagate them to the functional core, and if needed interpret them before.
- **Interaction** represents user input components such as a mouse, keyboard, magnetic pen, vocal recognition, etc.
- **Tools** are components that map interactions onto data processing or visualisation modification. For example, a mouse click will be translated into a rotation (view

manager) by the navigation tool but into a data modification (functional core) by a segmentation tool.

- **View Manager** manages all views and knows how data can be visualised on which type of view.

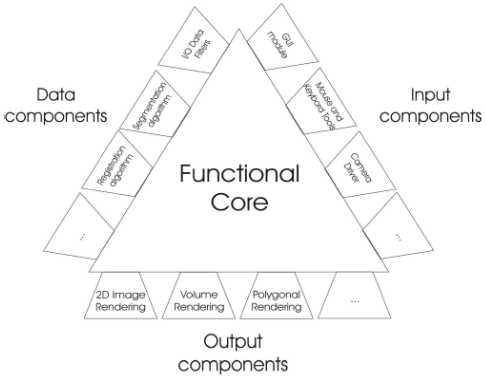


Figure 12. Medical Studio component architecture.

- **Renderers** are components that render specific data types onto specific views. An example is the rendering of an image on a 3D view with the raycasting algorithm.
-

Medical Studio architecture groups the data operators into data components and the visualisation operators into output components and adds input components (Figure 12). There are plans to refine this subdivision to add more control over the visualisation pipeline.

In order to assure cross-compatibility in terms of execution and development, Medical Studio is written in C++. The system also

supports easy integration of third-party software libraries for medical processing and rendering (e.g. ITK, VTK, and DCMTK).

4.2.2. GENERAL PURPOSE COMPONENTS

Medical Studio includes components for registration, segmentation and 3D reconstruction, and augmented visualisation. Thanks to the component-based architecture, when developing a new component in Medical Studio the programmer does not need to worry about implementing basic tools such as 2D and 3D rendering, view organisation and coordination, colour settings, transparency settings, or performing operations such as scaling, rotations, and zooming. All these functionalities as well as other components integrated into Medical Studio can easily share all the data required for their execution.

4.2.2.1. Multimodal Registration

Currently Medical Studio supports two kinds of multimodal registration: rigid and non-rigid.

1) Rigid Registration: Our rigid registration algorithm relies on the Insight Registration and Segmentation Toolkit. It makes use of a Simultaneous Perturbation Stochastic Approximation of the cost function gradient to seek the optimum. Another kind of rigid registration uses the surface-based algorithm. It minimises the mean square distance between the points representing the real object and the surface from segmented MRI images.

For that we applied the implementation of Saito's EDT (Euclidean Distance Transform) found in [2]. This method is re-used by other specialised components such as the Transcranial Magnetic Stimulation (TMS) application.

2) Non-rigid Registration: The use of the SPSA (Simultaneous Perturbation Stochastic Approximation) method has been investigated for optimising a large set of parameters characterising a non-rigid deformation. We use volumetric tetrahedral meshes as non-rigid deformation models. Their main advantage is their ability to deal with non-uniform sampling of the image domain and to allow multi-grid representation of the deformation.

The deformation is constrained by the linear elastic energy acting as a regularization term. This regularisation term can allow more flexibility in some regions than in others by giving different mechanical properties to elements in different regions.

4.2.2.2. Segmentation and 3D Reconstruction

This component allows the segmentation and appropriate labelling of anatomical structures for 3D reconstruction. As well as methods for manual segmentation where borders are drawn directly onto the raw image dataset, one of the classic methods for performing the task is *intensity based filtering* from MRI or CT dataset using the Marching Cubes algorithm [4] and acting in the same way as thresholding segmentation. The correct reconstruction also requires connectivity filtering to extract cells that share common points and satisfy a scalar threshold criterion.

Another algorithm implements automatic, *atlas-guided segmentation* [3], which is suitable for use in the presence of deformed anatomy caused by tumours and operates through a combination of rigid and non rigid registration components. The computed transformations map the atlas-segmented structures onto the subject volume. The component also includes level set segmentation [5] with active contour modelling for boundary object detection, letting an initial interface evolve towards the object boundary.

4.2.2.3. Augmented Visualisation

This component should support the real-time overlaying of computer generated graphics (i.e. virtual objects) on real objects and can be used to include pre-operative information in intra-operative video-based real scenes, thus enhancing the surgeon's perceptive capabilities. The component currently integrated in Medical Studio relies on stereovision to acquire the real scene and uses the surface-based registration component to perform an intra-subject registration between the real scene view and reference 3D mesh of the head skin.

4.2.3. PROCEDURE-SPECIFIC SCENARIOS

Procedure specific scenarios are examples of medical applications developed within the Medical Studio framework by making use of all the available generic components as well as the basic tools for view management, visualisation, and manipulation of images and volumes.

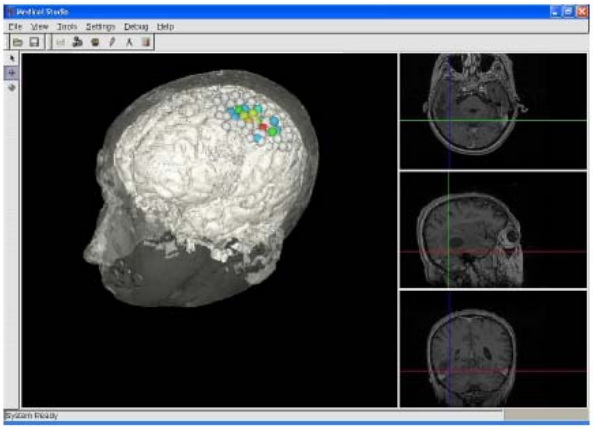


Figure 13. Visualisation of projection on brain surface from TMS session.

4.2.3.1. Transcranial Magnetic Stimulation

We have integrated into Medical Studio a scenario for registering and visualising in real time the results of Transcranial Magnetic Stimulations (TMS) in physical space on the corresponding anatomical locations in MR images of the brain.

The method proceeds in three main steps. Firstly, the patient's scalp is digitised in physical space with a magnetic-field digitiser, following a specific digitisation pattern. Secondly, a surface-based registration component, which minimises the mean square distance between those points and a segmented scalp surface extracted from the MRI, is used. Following this registration, the physician can follow the change in coil position in realtime through the visualisation interface and adjust the coil position to the desired anatomical location.

Thirdly, amplitude of motor evoked potentials can be projected onto the segmented brain in order to create functional brain maps (Figure 13).

4.2.3.2. Image-guided Surgery

This application allows the user to define the arrangement of a set of objects, including their multimodal annotations, in an augmented-reality composite scene and assures adequate rendering at run-time. Here the user is immersed in a full-scale physical environment viewed through a stereo-video head mounted display and interaction is based on hand-free user focus control.

In essence, a complex surgical procedure can be navigated visually with great precision by overlaying a colour-coded pre-operative plan on an image of the patient. It specifies details

such as the locations of incisions, areas to be avoided, and the diseased tissue. The added information can be anything from text to icons or colour-coding to three-dimensional surfaces.

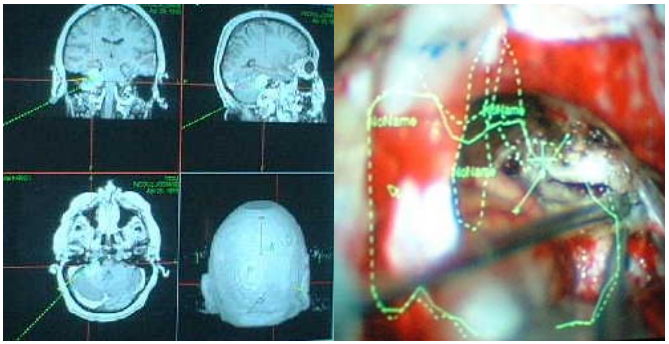


Figure 14 shows a microscope image-guided surgery system.

The method relies on the augmented visualisation component and can be used to define guidance support for surgical procedures. As well as identifying the elements related to the user's task, other decisions influence the way they are laid out in the real scene. These range from establishing their spatial and temporal relationships to setting viewing specifications and rendering parameters.

Three technical stages are necessary in order to achieve guidance for the surgeon in procedure execution:

1. In the **pre-operative planning**, the surgeon selects the structures of interest that must be segmented from the pre-operative images (CT, MRI, etc.) and may define additional information, such as annotations relevant to task execution (tests and *a priori* information about human anatomy contained in atlases or database).

2. **Registration and calibration phase,** where the user may define the correct spatial parameters to support augmented reality.

3. Finally, the virtual structures with additional annotations must be visualised and/or auditorily rendered during the surgical intervention, i.e**. intra-operative phase**. Figure 14 illustrates these tasks for a general scenario.

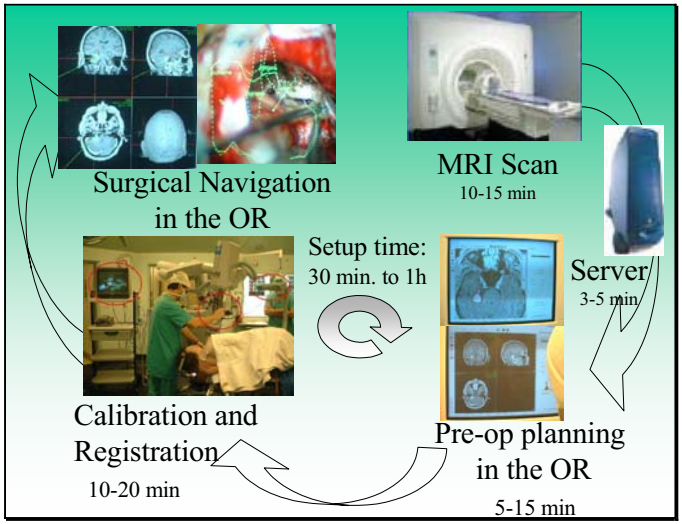


Figure 15. Image-guided surgery scenario where images acquired pre-operatively are registered to the patient images intra-operatively.

Human Factors in Intra-operative Enhanced Interaction

In user interfaces for critical tasks such as intra-operative support, cognitively adequate interaction is a key requirement to achieve usable and useful systems. This is especially true for augmented reality systems, where having multiple sources of information and two worlds of interaction (real and virtual)

involves making choices about what to attend to and when. To assess the user experience and effectiveness of enhanced multimodal interaction we consider how users perceive and evaluate an artifact's observable behaviour in order to infer about its state and plan and execute their actions. Usability criteria need to be extended or introduced to evaluate contextual visualisation and continuity of interaction during execution of parallel tasks in an intra-operative composite scene (including multimodal interactions).

4.3. OMICS Data Support

MedicalStudio is currently focused on organ and tissue image processing. In response to major developments over the last few years in molecular and cell biology, the platform will be extended to incorporate support for OMICS data and cover all the levels of the Biological Continuum (molecular biology information processing) This will be done for the particular case of radiotherapy in such a way that it will easily be re-used for other clinical cases. Furthermore, the software architecture will integrate new capabilities, including accessing, indexing, mining, processing, and visualising relations in large data volumes on the different levels of the continuum.

The data formats across the biological continuum need to be established and conversion standards investigated. Intuitive and efficient modes of visualisation (including a virtual microscope for tissues, see Figure 16) and interaction with the different types of information (biological structures, 2D/3D data) will be developed on the basis of ICARE. Two-handed interaction to manipulate DNA structures and zoom maps are candidate techniques for exploration and navigation within the large information space.

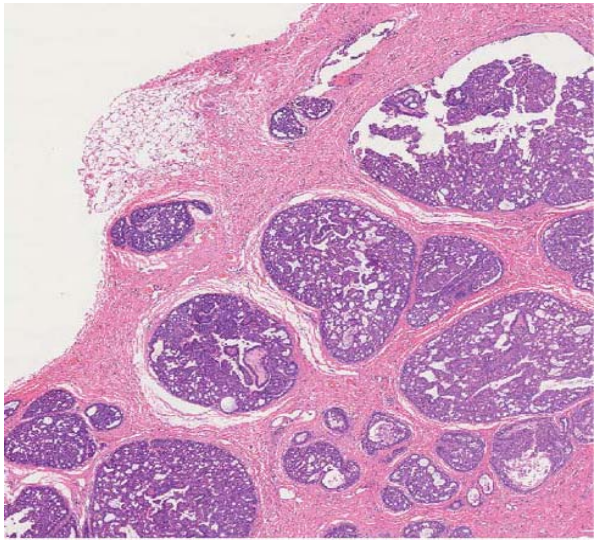


Figure 16. Virtual slide viewing in the atlas of breast histopathology

Treatment planning tools for radiotherapy will include anatomical and propagation atlases for considering dose planning constraints in healthy tissues and tumour, use of a patient genotypic profile to predict the tumour radio resistance profile, etc.

Computer-aided anatomical image segmentation will allow optimisation of dose delivery to the tumour while preserving healthy tissue and organs. As an example, prior models regarding the deformations of organs around the tumour and respiratory movement for lung images will be included in the atlas-based segmentation framework.

4.4. References

Bouchet, J., Nigay, L., (2004) “Icare: A component-based approach for the design and development of multimodal interfaces,” in Proc. of 11th International Conference on Human-Computer Interaction HCI International. Vienna: Lawrence Erlbaum Associates.

Cuisenaire, O. (1999) “Distance transformation, fast algorithms and applications to medical image processing,” Ph.D. dissertation, Universite catholique de Louvain, Louvain-la-Neuve, Belgium.

D Haese, P., du Bois d Aische, A., Merchan, T., Macq, B., Li, R., Dawant, B., (2003) “Automatic segmentation of brain structures for radiation therapy planning,” in Proc. of the SPIE (vol.5032): Image Processing, San Diego, USA, pp. 517–526.

Lorensen, W. E., Cline, H. E., (1987) “Marching cubes, a high resolution 3d surface construction algorithm,” Computer Graphics, vol. 21, no. 4, pp. 163–169.

Q. Noirhomme, M. Ferrant, Y. Vandermeeren, E. Olivier, B. Macq, and O. Cuisenaire, Real-Time registration and visualisation of transcranial magnetic stimulation with 3D MR Images, IEEE Transactions on Biomedical Engineering, Vol. 51, No 11, pp. 1994-2005, 2004.

Osher, S., Sethian, J. A., (1988) “Fronts propagating with curvature dependent speed: a: algorithms based on hamilton–jacobi formulations,” Journal of Computational Physics, vol. 79, pp. 12–49.

Chapter 5: Building Usable Multimodal NIS

Niels Ole Bernsen
& Laila Dybkjær (NISlab, Denmark)

5.1. Introduction

This chapter describes the SIMILAR Special Interest Group (SIG) on usability's approach to research on usability evaluation of multimodal and natural interactive systems (NIS).

Before presenting our plan for the chapter, it seems useful to take a closer look at the key terms of the preceding paragraph. *Usability evaluation* is part of systems and component evaluation more generally. However, usability evaluation is not evaluation of yet another system component. Rather, system usability is affected by the performance of all or most system components as well as their integration into the system, which is why usability evaluation cannot ignore the results of more technology-oriented evaluation. Usability evaluation is also part of the remit of the field of human-computer interaction (HCI) which has been an active research area for more than 30 years. However, until much more recently, say, the mid-1990s, HCI research (i) was dominated by research on GUIs (graphical user interfaces) and (ii) suffered from the fact that many HCI researchers did not form part of software development teams. Research on interaction

and usability of other, non-GUI technologies, such as interactive speech systems, has been carried out more or less independently from the HCI community by developers of interactive speech systems. Research on the usability of natural and multimodal interaction had to await the emergence of the corresponding technologies in the past 5-10 years.

A *natural interactive system* interacts with users by means of one or several of the modes that humans use to communicate with one another, such as speech, gestures, facial expression, body posture, object manipulation as part of communication, hand-writing, possibly typing, etc. A *multimodal system* is a system which uses more than one mode of information exchange with the users. Thus, multimodal systems employ several individual, or *unimodal,* modalities, either as input modalities, as output modalities, or both. It follows that natural interactive systems can be made *more* naturally interactive through relevant multimodality and that multimodal systems are not necessarily naturally interactive.

It is worth keeping in mind that the GUI system paradigm is itself multimodal, taking several kinds of haptic input provided by *input devices,* such as keyboard and mouse, and outputting graphics in many different modalities. Yet, arguably, in the past ten years or so, the world of multimodal systems has been augmented with literally scores of new input/output modality combinations, reducing past results on the usability of GUI interfaces to a small fraction of the complexity facing us today. To some extent, the usability of individual systems representing many of those modality combinations has been studied already. However, the sheer complexity of the task demands more systematic approaches than single-system usability evaluation, if, indeed, any more systematic approaches are possible. This question is the point of departure for the SIMILAR usability SIG which, furthermore, aims to address the question, for a start, at least, based on the specific premises of the SIMILAR network (see Section 3).

In the following, we briefly outline the state of –the art in usability evaluation in multimodal and natural interactive systems (Section 2). Section 3 presents the SIMILAR Usability SIG's objectives and the approach taken to address these issues. The approach includes as one of its first steps a common application description structure which is described and exemplified in Section 4. Section 5 discusses next steps in the Usability SIG's work.

5.2. State of the Art in Brief

Apart from GUI applications, which are inherently multimodal, research has for many years concentrated on unimodal systems. In recent years, research systems have been moving towards combining several input and/or output modalities, as in talking heads or embodied conversational agents, in-car applications using spoken dialogue and a small display, games using computer vision input and graphics output; and many others.

This recent trend has generated a need for knowledge of how to evaluate the usability of multimodal systems. In many respects this remains an open research issue. We are not necessarily starting from scratch, though, since it would seem obvious to draw on methods and criteria from usability evaluation of unimodal systems to the extent that they are transferable to the multimodal context.

However, even as regards unimodal systems there is a major gap in our usability evaluation knowledge. This gap concerns what usability actually is and what exactly makes a user like, or accept, a system. We know that there are several factors contributing to user satisfaction but we do not know them all or the extent to which each of them contributes. Moreover, the importance of each factor may differ across users and user groups.

When addressing multimodal systems a new main challenge is to find criteria for evaluating the combinatorial contribution to

usability and user satisfaction of the input/output modalities involved. For the moment, exploration of how well different modalities work in combination and of their effects on users is often carried out via comparative studies of users interacting with different systems. However, the findings of such studies are not generalisable to any larger extent, which means that for new applications new studies must usually be made. An alternative, theory-based approach is to continue to develop heuristics based on Modality Theory [Bernsen 2002].

In addition to the modality combination problem, there is also the continued proliferation of new system types and the increasing sophistication of systems, whatever their modalities, both of which factors continue to demand new usability evaluation metrics. For example, systems may be operated in mobile environments and not only in a static environment. Other recent system type innovations include systems for education, edutainment, and entertainment. As regards increased sophistication, there are now systems that explore the inclusion of on-line user modelling to provide more flexible and adaptive interaction behaviour. Some systems aim to recognise the user's emotional state and/or to exhibit emotional states of their own, in both cases in order to provide more appropriate and natural system reactions. User preferences and priorities raise new issues in such systems.

A number of handheld and other mobile devices that allow multimodal applications have become available. One example is mobile phones that allow spoken as well as key-pad input and PDAs that allow pen-based input in addition to spoken and keypad input. Another example is in-car applications. Mobile systems raise several evaluation issues that have not been fully solved, including how (not) to use, and when (not) to use which modalities and which input/output devices, and for which purposes (not) to use location awareness and situation awareness.

Usability evaluation often includes application of the three ISO-recommended (International Standardization Organisation,

www.iso.org) usability parameters, i.e., effectiveness, often measured as task success rate; efficiency, often measured as time to task completion; and user satisfaction, often evaluated based on a questionnaire (see. ISO 9241-11). Even as regards these basic approaches, however, it may be noted that some of the new system types are not task-oriented at all. Entertainment systems are a case in point. For such systems, arguably, considerations of effectiveness and efficiency – at least in the traditional sense - are simply irrelevant, whereas new usability criteria, such as some new form of interaction success, entertainment quality, and interaction naturalness, come to the forefront when evaluating usability.

On-line user modelling is receiving increasing attention for several reasons. Users of mobile devices that are usually personal belongings may benefit from functionalities that build up knowledge of the individual user. Generic user modelling may also be useful. For instance, novice users could receive more extensive interaction guidance and users who repeatedly make particular types of mistake could be helped by explicit advice or by adaptation of the interaction structure. Some key evaluation questions regarding on-line user modelling concern: (i) if the user modelling functionality is technically feasible in the first place and (ii) whether it will be of benefit rather than a nuisance to the majority of the application's users. For instance, even if the system has enough information on an individual user, adaptation may fail because of too primitive update algorithms or insufficient information for the user about when the user model has been applied.

Not only recognition of users' emotional states but also systems' expression of emotion is an active research area. Usability evaluation must consider what positive and negative impacts emotion modelling has on users.

User preferences can make life hard for the developer as they may contradict what is empirically the most efficient solution.

Some users may, e.g., prefer pen-based input to spoken input or keypad-based input to spoken input simply because they feel more familiar with GUI-style interfaces. Depending on the target user group(s), alternative modalities may be needed because it is likely that each of them will be preferred by some significant fraction of the user population. This is just one reason why user involvement from early on in the development process is recommended and on-line user modelling appears attractive. Some preferences we can design for, such as modality preferences. Others, however, are hard to cope with. Thus, some users may prioritise speed or economical benefits, while others prioritise human contact. The question is whether we can build systems with a usability profile that will make the latter users change their priorities, and exactly which usability issues must be resolved to do so.

In brief, there seems to be a broad need for usability evaluation that can help us find out how users perceive new kinds of multimodal and natural interactive systems and how well users perform with them, possibly compared with other types of system. There is a strong wish in the field to find ways in which usability and user satisfaction can be correlated with technical aspects in order for the former to be derived from the latter. We do not have methods today that can reliably predict how well users will receive a particular system. We just know that a technically optimal system is not enough to produce user satisfaction. Regarding modality appropriateness, which is a central issue in multimodal SDSs, modality theory may be a promising and powerful approach to usability evaluation of modalities at an early stage. However, user tests of the actual design will still be needed, just as for unimodal systems. For an overview of usability evaluation of multimodal systems, in particular systems involving speech, see [Dybkjær et al. 2004].

5.3. The SIMILAR Usability SIG Approach

The SIMILAR Network has particular properties that should be taken into account by the Usability SIG. Thus, (i) SIMILAR focuses on natural interaction rather than on multimodal interfaces more generally ("... taskforce creating human-machine interfaces SIMILAR to human-human communication"). Yet, (ii) SIMILAR includes research issues, such as (non-invasive) brain process interpretation, which clearly seem to go beyond human-human communication. To the Usability SIG, the SIMILAR community is already granting access to a selection of innovative multimodal natural interactive technologies, limited only, to some extent, by (iii) the network's focus on three particular application areas, i.e., medical, disability, and edutainment.

Finally, (iv) SIMILAR has a preponderance of signal processing researchers, which places interesting demands on the Usability SIG because, from a traditional point of view, signal processing does not form part of software engineering at all but, rather, addresses a large class of basic technologies for potential take-up by software engineers and system developers. From a usability point of view, important questions may arise in this context about how to evaluate the usability of new and promising signal processing algorithms that, as is often the case, are "looking for applications" rather than being applications by themselves. This is not usability evaluation in any standard sense of the term.

5.3.1. General objectives in brief

At the time of writing we are only about half a year into the SIMILAR project. The objectives for the SIMILAR Usability SIG for the first 18 months are to

- establish the group of SIMILAR members who will contribute to the Usability SIG;
- review relevant literature on usability evaluation of natural interactive and multimodal applications;
- establish a pool of accessible natural interactive and multimodal applications (exemplars) developed within SIMILAR that can be analysed in depth from the point of view of usability evaluation;
- create a template-like analytical structure for obtaining a detailed description of how each application has been evaluated regarding usability;
- use the structure to describe current practice for the available pool of applications, possibly adding resources permitting, the results of new usability evaluation exercises with respect to selected systems;
- based on the above, develop a first outline of a best practice framework and guidelines for the evaluation of system and component usability in the field of natural interactivity and multimodality.

The best practice outline will necessarily be a very preliminary outline, given the effort required and time available during the first 18 months as well as the size and complexity of the field. Thus, the main focus during the last 30 months of SIMILAR will be to consolidate iteratively and enlarge the coverage of the best practice framework. Scope consolidation and enlargement will be achieved partly through analysis of the issues arising and partly through analysis of additional exemplars from SIMILAR partners and colleagues outside SIMILAR.

In Section 4 we present the exemplar description structure and exemplify its use. Prior to that, we would like to discuss in more general terms our approach to outlining current practice (Section 3.2).

5.3.2. A bottom-up approach

The approach we have decided to use in the SIMILAR Usability SIG is inspired from, and partially builds upon, the DISC project (Spoken Language Dialogue Systems and Components: Best practice in development and evaluation, www.disc2.dk) in which academic and industrial partners investigated current and best practice in the development and evaluation of spoken dialogue systems and their components. The DISC current practice approach was to (a) analyse a broad range of spoken dialogue systems and components and (b) map out their respective development and evaluation processes. In order to capture current practice adequately and overcome various problems primarily relating to the insufficient and not-easily-comparable information provided for individual systems and components, a common scheme was developed.

This scheme was applied to the analysis of about 25 exemplars, i.e., systems and components to which the project partners provided access . Each exemplar was analysed independently by two different project partners, yielding 50 confidential internal reports. For each component level and the system level a synthesis description was made based on the relevant exemplar descriptions. Each synthesis description abstracted from individual component- or system-specific observations and presented the range of practical approaches followed in the development and evaluation of systems or components. Based on normative analysis of the current practice descriptions a draft best practice was then established.

By contrast to the SIMILAR Usability SIG, and apart from a single multimodal system present in the DISC exemplar pool, DISC focused on task-oriented unimodal spoken dialogue systems, and DISC considered not only– technical as well as usability - evaluation but also the entire development process and which features to include when building a spoken dialogue system. In the Usability SIG we shall look solely at usability

evaluation, which is a simplification compared with DISC. However, the range of systems to be included will be much larger, more varied, and more complex than in DISC.

As in DISC, we have established a first pool of systems kindly made available for usability evaluation research by the SIG participants. The exemplar pool includes a total of seven applications, including two entertainment/edutainment systems, three surgery/operating theatre systems, a museum application, and a training system for the blind. An in-car application may be added later. Individually, the systems are very different and, collectively, their properties go far beyond those addressed in the DISC scheme.

To solve this problem, we are drawing upon the approach used in the MATE (Multilevel Annotation Tools Engineering, mate.nis.sdu.dk) and ISLE (International Standards for Language Engineering, isle.nis.sdu.dk) projects to collect information about, and subsequently describe, a wide variety of different natural interactivity data resources, annotation schemes, and annotation tools. In both projects, common description structures were developed and applied to the collected information about each data resource, annotation scheme, and annotation tool.

This turned out to work rather well. We have therefore adopted a similar approach in the Usability SIG, establishing a common description structure that will be described in more detail in Section 4. Once we have collected the corresponding information about all SIMILAR exemplars, we shall look into the problem of establishing and refining a scheme for capturing current practice descriptions of usability issues in the systems.

5.4. Application Description Structure

In the following section we introduce (Section 4.1) and exemplify (Section 4.2) the common structure used for exemplar descriptions in the SIMILAR Usability SIG.

5.4.1. Introduction

The purposes of having a common application description structure are to ensure (i) presentation, at a common level of detail, of all applications to be analysed and evaluated in the SIMILAR Usability SIG, and (ii) a minimum of information on each application, subject to additional information gathering when required for usability evaluation purposes.

The precise entries of the common application description structure are shown in Section 4.2. It is quite possible that some of the entries can be filled in only tentatively, if at all, for a particular exemplar, such as when the system is still under development and usability evaluation is still ongoing.

Reports may be classified confidential if a contributor so wishes. This means that their contents will become public only at the higher, more abstract level at which all references to the properties of particular systems will be removed. Report contributors will of course have the opportunity to check, prior to publication, that confidential information has been removed at that higher level of presentation.

5.4.2. Application description

In the following we exemplify the entries in our application description structure. The examples draw upon four internal Usability SIG documents, i.e. [Bernsen and Dybkjær 2004] for the NICE Hans Christian Andersen (HCA) example, [Trevisan et al. 2004] for the Image-guided Neurosurgery example, [Gómez et al. 2004] for the SINERGIA laparoscopy virtual simulator, and [Berti et al. 2004] for the Portable Cicero museum application.

5.4.2.1. Purpose of the application

Insert the goals of the development project. These will typically be to demonstrate, or deliver to customers, certain functionalities, often adding usability descriptors and aims regarding the user's appreciation of the application. Other goal parameters may include price and quality descriptors, target users, use settings, etc., informally described and further detailed below.

Example: The NICE HCA system's main goal is to demonstrate natural human-system interaction for edutainment, in particular involving children and adolescents, by developing natural, fun, and experientially rich communication between humans and embodied historical and literary characters.

Example: Image guided surgery is a type of computer-assisted surgery that uses advanced three-dimensional visualisation techniques to provide the surgeon with a wealth of valuable information not normally available in the operating theatre.

5.4.2.2. Input modalities

Describe the way(s) in which the user inputs information to the system, either using the terminology of Modality Theory [Bernsen 2002] or using informal descriptions, such as "GUI-style input augmented with ...", possibly referring to input *devices* rather than modalities, or referring to both modalities and devices. For instance, we all understand what standard mouse input is even if we do not realise that the mouse is a simple haptic input code device.

Example: Spontaneous English speech and 2D gestures via mouse or touch screen (HCA system).

Example: Pen and infrared waves (museum application).

5.4.2.3. Output modalities

Describe the way(s) in which the system outputs information to the user (see Section 4.2.2).

Example: 3D animated, life-like embodied HCA communicates with the user through English conversational speech, gestures, facial expressions, body movement, and action (HCA system).

Example: Speech, video, map, and text (museum application).

5.4.2.4. Target user group(s)

Arguably, no interactive application can be meaningfully targeted at all users/ Just think of the different user properties, such as the language(s) that they speak, their culture, their age, their educational and professional background, their interests, their mastery of the human senses, their mastery of their body, etc. It follows that any application must target (a) specific user group(s).

Example: *T*arget users are 10 to –18-year-old children and teenagers (HCA system). Basically, the system requires no training in order to start using it.

Example: surgeons (Image-guided Neurosurgery).

5.4.2.5. Physical use environment

Despite application classifiers like "ubiquitous computing" and "ambient intelligence", most applications are intended for use in or across specific environments.

Example: *T*he primary use setting of the HCA system is museums and other public locations.

Example: Image-guided Neurosurgery is meant for use in operation theatres.

5.4.2.6. Which domain does the application cover?

Enter what the application "is about", i.e. the general domains of information, action, or otherwise, that the application addresses.

*Example: T*he general domain of the HCA system is a combination of education and entertainment. More specifically, the system allows users to have a conversation with an HCA about his life and fairytales, himself and his study, as well as about the user, games, and technical inventions.

Example: Image-guided Neurosurgery is within the medical domain.

5.4.2.7. Which tasks (if any) does the application solve?

A task is a far more specific entity than a domain. Most applications are aimed at enabling the user to do a, or some, more or less specific task(s). Some applications, however, are not task-oriented at all in any clear sense of this term. The HCA system is an example of the latter.

*Example: t*he system is mainly aimed to support two main tasks: to help users orient within the museum, and to provide them with multimedia information at different abstraction levels (museum, section, physical environment, single work) (museum application).

*Example: t*he SINERGIA laparoscopy virtual simulator focuses on simulating different basic steps of Nissen founduplicature, including grasping and pulling, cutting, dissection, and suture.

5.4.2.8. Is the application free or what is the price?

This question is obviously important to anyone interested in the application. If the application is "free", the user might want to know if this means that it is open source or if a free executable is available. Moreover, the user will want to know how to get ahold of the system.

Example: This research prototype application is not free, nor does it have a price. If someone wishes us to, e.g., port the application to a different language or even replace the HCA with a different character, please contact us and we shall estimate the cost (HCA system).

Example: Image-guided Neurosurgery is a proprietary system. No price information is available.

5.4.2.9. If not free, is a demo available?

Among other things, demos include short-time test licenses, reduced-capability system versions, simple demos, such as sound or video recordings of human-system interactions, etc.

Example: A small demo video is available at www.niceproject.com/ about/ (HCA system).

Example: The Portable Cicero museum application is currently available for all Carrara Marble Museum visitors. Its use is free. Further information is available at the following URL: http://giove.cnuce.cnr.it/cicero.html.

5.4.3. Technical issues

5.4.3.1. Platform(s) (operating system(s))

This is crucial information on any system. In many cases, the information should be supplemented by information on specific

non-standard software needed to run the application, including APIs. For research prototypes in particular, it is important to describe which platform compatibilities have actually been tested. This also applies to "platform-independent" software. Any other "exotic" information needed to run the software should be listed here as well, such as specific platform settings that are required but may not be intuitively obvious.

*Example: T*he HCA system runs on a Windows 2000 platform. It has not been tested on any other platform and for the moment there are no plans for testing the system on other platforms.

*Example: T*he SINERGIA laparoscopy virtual simulator runs on a Windows NT/2000/XP platform and makes use of two personal computers (PCs) in a client-server architecture. One PC runs the simulation software (collision detection, biomechanical model, and visual rendering) and the other controls the haptic device, sensing the position of the virtual tools and rendering the interaction forces to the user.

5.4.3.2. Hardware requirements

In principle, this information should be provided for any application, however small and limited in its requirements. Potential users of the application should receive sufficient help rather than having to do under-informed guesswork.

Example: Running the system requires a powerful computer with 500-1000 Mb RAM and a good graphics card, such as G-Force 4 (HCA system).

Example: One special piece of hardware is needed for the simulator, namely, the haptic device. Our choice has been the Laparoscopic Surgical Workstation (Immersion Corp., San Jose, U.S.A.). The two PCs required do not have special requirements, but it is advisable to have as much computational power as possible in order to increase real-time performance. It has been observed that if the simulator is run on a PC Pentium III, 450 MHz, with 256 RAM memory, visual and haptic update rates of

15 and 500 Hz can be reached (SINERGIA laparoscopy virtual simulator).

5.4.3.3. Implementation language(s)

This information backs up the potentially complex information provided in Sections 4.3.1 and 4.3.2. In the case of highly complex systems, shortcuts are permitted.

*Example: T*he implementation languages used in the HCA system are mostly Java, C++, and Sicstus Prolog. The HCA system is an example of a complex system for which the need to provide detailed module-per-module programming language information is not obvious.

*Example: T*he SINERGIA laparoscopy virtual simulator has been written in C++. For development the Microsoft C++ environment has been used.

5.4.3.4. Architecture

High-level knowledge of the system architecture is important even in the case of usability evaluation. This knowledge helps understand possible usability shortcomings and supports the asking of additional questions when such shortcomings have been discovered. Preferably, the architecture should be presented via an annotated high-level architecture diagram. The annotations should describe the individual modules, their origin, and the overall information flow.

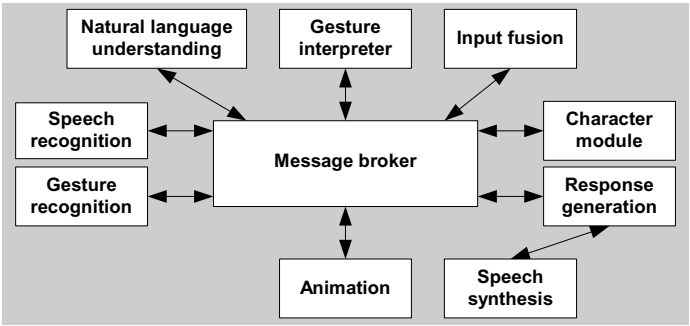


Figure 17: HCA architecture

Example: The diagram above shows the overall architecture for the HCA system. Explanations of components are not included in this brief example.

Example: The SINERGIA laparoscopy virtual simulator is composed of four main modules controlled by the main programme (reference to architecture figure not included in the brief example excerpted here):

- biomechanical model: calculates deformation and the behaviour of the organ in the virtual scene;
- collision module: calculates the interaction between the virtual models;
- visual motor: represents the geometry in the visual device (screen);
- haptic motor: reads the positions of the haptic device and returns the haptic forces to the user.

5.4.4. Functionality

5.4.4.1. Which functionality does the application offer?

System functionality is closely related to the system's purpose and task(s). System evaluators are likely to address the

described system functionality by asking if, e.g., the functionality is adequate for the system's purpose and the tasks supported by the system. If, as is often the case in research prototypes, the existing system functionality is deemed to fall short of the desired functionality, it is important to describe the functionality that is deemed missing so as to spare the usability evaluators the effort of pointing this out.

The properties of system functionality ("what the system can do for you") and usability ("how usable the system's functionality is") are analytically distinct but closely related. A functionally adequate system may be partly or wholly impossible to use due to its inadequate user interface. Conversely, a functionally inadequate system may become popular among users due to its intuitive user interface. Listing the system's functionality is prerequisite to enabling the usability evaluator to judge if the functionality is adequate for the system's task(s), or otherwise, and whether the existing functionality is actually usable. An inadequate user interface can effectively hide much useful functionality from the user, except for users who study the manual carefully or receive substantial training.

Example: The SINERGIA laparoscopy virtual simulator is task-oriented, i.e., there are different particular tasks which the user is meant to perform with the system. In our first prototype there will be four surgical tasks: grasping and pulling, cutting, dissection, and suture.

Example: The Portable Cicero museum application provides the users with information about the artworks located in the marble museum using the devices' multimedia capabilities and taking the user's position into account .

5.4.4.2. Description of each main functionality

For complex systems, it is sufficient, at this point, to list the major system functionalities, referring to any supporting documentation for more detail.

Example: The SINERGIA laparoscopy virtual simulator first prototype includes the following four surgical tasks:

- Grasping and pulling: The user has to orient the surgical tools to grasp different objects and pull them until tearing occurs. With this basic task users familiarise themselves with different tissue consistencies and tearing thresholds, learning how to interpret visual and haptic information.
- Cutting: The user has to cut a surface through a line drawn on it. Metrics of precision, time, and economy of movement are recorded, which allows the user to make an auto-evaluation of the task, providing constructive feedback.
- Dissection: The user has to dissect two virtual layers of a virtual stomach. Metrics of error, time, and economy of movements are recorded.
- Suture: The user has to perform a virtual suture on a flat surface.

Example: During the visit the user can perform the following tasks:

- Orientation within the museum. For this purpose three levels of spatial information are provided, namely, a museum map, a section map, and, for each physical environment composing the section, a map with icons indicating the main pieces of work available in the room and their locations.
- Control the user interface, for example, to change the audio commentaries' volume , to stop and start the commentaries, and to move through the various levels of detail of the museum information available;

- Access museum information. This is also provided at different abstraction levels (museum, section, physical environment, single work).
- Path Finder method allows visitors to find the location of an artwork they are interested in by suggesting the path to reach it starting from the room they are in. The result of the request is a map highlighting the section where the user currently is, the section where the artwork is located, and the path that the user has to follow to find the artwork. The next figure shows an example of the result of a user invoking this feature and interacting with the system.

5.4.5. Interface and usability

5.4.5.1. Design description

Given the system functionality descriptions in Sections 4.4.1-2, this section describes the usability aims, heuristics, guidelines, standards, underlying theory, or otherwise, or the lack of them, which were adopted in designing the user interface. The tentative or consolidated nature of the user interface should be described as well.

Example: In the SINERGIA laparoscopy virtual simulator the user interacts through two sensory channels at the same time: the visual and haptic channels. When he moves a virtual tool with the haptic device, this tool is moved in the monitor and interacts with the virtual organs, and he feels the interaction forces. All this is performed several times a second (visual update rate: 15Hz; haptic update rate: 500Hz).

Example: Designing an application for a PDA should take into account the specific features of this type of device, as it provides a broader range of interaction techniques than current mobile phones. The possibilities are similar to those of desktop systems, with two main differences, *i.e.*, the limitation of the screen

resolution and the possibility of using it on the go (museum application, abbreviated description).

5.4.5.2. Which user skills (if any) are assumed?

This entry should mention any in-context, non-trivial skill requirements that the user is assumed to satisfy. At this point in the application description, we already have plenty of context. Thus, if the application is a statistics package, for example, the task is to use the package for statistical purposes and the standard assumption will be that the user is familiar with the principles of statistics. However, if the statistics package did not assume user knowledge of statistics, this would be contextually non-trivial information. If no particular skills are required, this should be stated as well.

Example: Very good knowledge of the system is required (Image Guided Neurosurgery). Moreover, the user must be educated as a surgeon.

Example: There are no skills assumed for the user except that users are supposed to have some knowledge of surgical tools, terminology, and procedures. The application starts with a menu in which the user selects the training task he wants to practice, and when the task is finished, a report is provided to the user to provide constructive feedback about his performance. (SINERGIA laparoscopy virtual simulator).

5.4.5.3. Is training foreseen? Is there a manual?

A walk-up-and-use system is a system designed for ordinary users who should be able to use the system without any training or manual consultation. It is important to describe any non-trivial learning (or training) requirements imposed by the application. If there is a user's manual, information on how to access the manual should be provided.

Example: The first HCA system prototype requires only that the user know how to change camera angles using a function key, control HCA's locomotion using the arrow keys, and use a touch screen, if available (otherwise, the mouse may be used). There is no manual for the system.

Example: No particular skills or experience are required. There is no manual (museum application).

5.4.5.4. Illustrative examples of interface use

User interface illustrations are important for early inspection of the look-and-feel of the system and prior to trying out the system first-hand. Illustrations especially concern static or dynamic graphical user interfaces, haptic input or output devices, and the like. For spoken/acoustic interfaces or interface parts, transcriptions of example interactions, or sound files, are helpful.

Example: Use of the HCA system in the HCA museum in Odense Denmark.

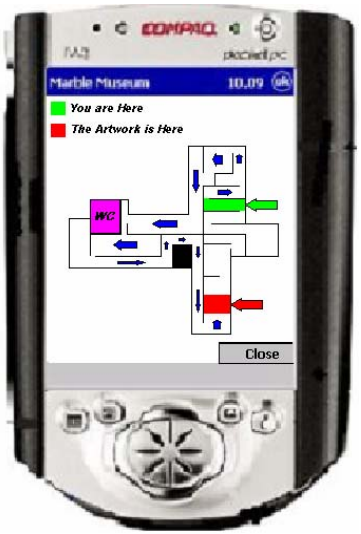


Figure 18: PDA showing museum map (museum application)

5.4.5.5. Assessment of functionality and interface

This entry should provide a general assessment of the system's functionality and interface, possibly referring to the functionality descriptions in Sections 4.4.1-2. The assessment may reflect the developers' anticipations only, but it may also reflect general lessons learned from user tests performed with the system.

Example: The first HCA system prototype of January 2004 is still incomplete and primitive in several basic respects: speech recognition is not integrated yet, graphics rendering is still primitive and seeing the large number and the timing of non-verbal behaviour primitives, the spoken conversation is too inflexible; the speech synthesis must be also improved.

5.4.6. Evaluation

5.4.6.1. Who/how many have used the application so far?

For products, this question may be unanswerable, of course. For research prototypes, on the other hand, the information provided is likely to be important. For instance, if only the developers have used the application, this is important information. Also, for research prototypes, the test users who have worked with a more or less simulated system version should be described together with the nature and setup of the simulation.

Example: Eighteen kids and teenagers have used the first HCA prototype in a controlled lab test.

Example: Thirty-five museum visitors have used the system and subsequently filled in a questionnaire (museum application).

5.4.6.2. How has the application been usability tested?

If one or several user tests have been made with the system, it is important to describe the test protocol and evaluation criteria applied, possibly referring to additional accessible information.

Example: Museum visitors were given a PDA with the museum application installed and after their visit to the museum they filled in a questionnaire. The goal of the test was to understand to what extent the application provides a valid support from various viewpoints, *i.e.*, quantity and quality of information provided, modality of presentation, interaction with infrared devices, and capacity to help users orient in the museum.

5.4.6.3. Which evaluation results are available so far?

This section should describe any evaluation results obtained (including references to where they are documented). If the results are complex and detailed, a general description is sufficient together with references to additional accessible information. If future user tests are being planned, this should be explained.

Example: The answers collected from the 18 users who participated in the user test were, encouraging, even surprisingly so. Overall, the users felt that the technology was on the right track and represented a first glimpse of entirely new spoken computer game technology that could significantly improve the entertainment and educational value of computer games as well as attracting a new group of users who have not been so interested in traditional computer games. More information about the evaluation of the interview data can be found in [Bernsen and Dybkjær 2004].

5.4.7. Conclusions

5.4.7.1. General assessment

This entry should provide general lessons learnt so far, if any, on the system's usability. If additional usability evaluation of the system is required, this should be stated, preferably with details on the aspects needing further investigation. It is useful to mention ongoing work on analysing already gathered user test data as well.

Example: Basically we are interested in measuring the user’s interaction with the system during the surgical intervention. Then the continuity of task and interaction are very important points for the success of the application. A new multimodal/augmented

system (with alternative interactions) is being developed (Image Guided Neurosurgery).

Example: The surgical simulator of the SINERGIA network is being developed for the laparoscopic training of surgeons. The simulator will be inserted in the training curriculum of the MISC of Cáceres (Spain), where the evaluation studies will be performed.

5.4.8. References

5.4.8.1. More information about the application

It is useful not only to provide references but also to explain, for each reference, the information that it contributes. For confidential references, please state what may be disclosed, and possibly in which way.

Example:

Bernsen, N.O. and Dybkjær, L.: Evaluation of Spoken Multimodal Conversation. Proceedings of the International Conference on Multimodal Interfaces (ICMI), 2004 (in press).

Paper describing HCA PT1, focusing on multimodal conversation and user test evaluation results on multimodal conversation.

5.5. Next Steps

At the time of writing, we have almost finished writing the system descriptions following the common structure presented in Section 4. The next step will concentrate on the evaluation description given per application to see how far we can get in extracting and synthesizing a current practice scheme for usability evaluation.

This is likely to be an iterative process requiring more detailed information about the evaluation criteria and methods used than has been made available in the descriptions. It may also turn out

that we need to include additional applications to consolidate the scheme. When we have a reasonable current practice description we shall try on that basis to distil a draft best practice framework from it.

5.6. References

Bernsen, N. O.: Multimodality in Language and Speech Systems - from Theory to Design Support Tool. In Granström, B., House, D., and Karlsson, I. (Eds.): Multimodality in Language and Speech Systems, Kluwer Academic Publishers, Dordrecht, 2002, 93-148.

Bernsen, N. O. and Dybkjær, L.: Description of the NICE Hans Christian Andersen System. NISLab, University of Southern Denmark, Internal SIMILAR Report, 2004.

Berti, S., Paternò, F. and Santoro, C.: Description of the Portable Cicero Application. HIIS Laboratory, ISTI – CNR, Internal SIMILAR Report, 2004.

Dybkjær, L., Bernsen, N. O. and Minker, W.: Evaluation and Usability of Multimodal Spoken Language Dialogue. Speech Communication, Vol. 43/1-2, Elsevier, 2004, 33-54.

Gómez, E., Lamata, P. and Rodríguez, S.: Description of the SINERGIA Laparoscopy Virtual Simulator. GBT, Polytechnic University of Madrid, Internal SIMILAR Report, 2004.

Trevisan, D., Gemo, M., Nicolas, V., Vanderdonckt, J. and Macq, B: Description of an Image Guided Application Using Contextual Focus Driven Interaction. UCL-TELE and BCHI, Université Catholique de Louvain, Internal SIMILAR Report, 2004.

DISC project: www.disc2.dk

ISLE project: isle.nis.sdu.dk

ISO: www.iso.org

MATE project: mate.nis.sdu.dk

Chapter 6: Multimodal Interfaces for People with Disabilities

Dimitrios Tzovaras (ITI-CERTH, Greece)

6.1. Introduction

The rapid development of Information and Communication Technologies (ICT) and emergence of the Information Society (IS) create numerous new opportunities for interpersonal communication, education, vocational training, employment, entertainment, etc. They also entail, however, a number of risks, mainly concerning the creation of a society of “haves” and “have-nots” with regard to access to the applications and services offered by the IS, which can significantly compromise society’s cohesion.

People with disabilities constitute one of the most important categories of citizens at risk in this context. Although the emergence of the IS creates tremendous opportunities for the improvement of life and independent living of the disabled, it can be argued that people with disabilities have not yet realised the full potential of the IS. In particular, most applications and services are designed and developed for the “average”, able-

bodied user, and do not take into account the abilities, requirements, preferences, skills, interests, etc, of people with disabilities, and, as a result, “mainstream” technology is not usually accessible to the disabled.

In this context, the major challenge for activities in the area of multimodal interfaces for the disabled within SIMILAR is to set up a strong research group that will join efforts in the development of innovative interfaces integrating multiple modalities (text, speech, vision, haptics, emotions, 3D, etc.) for the communication and training of people with missing or impaired interaction channels, *e.g.*, the blind, deaf, quadriplegics, etc. Also, the challenge is to provide a test bed for the development and assessment of multimodal interface for people with disabilities that can provide a solid basis for further research and significantly advance the state –of –the art in the area.

SIMILAR will focus on developing novel algorithms for the registration and integration of multimodal data, to provide new services to people with disabilities. In the specific activity of SIMILAR we shall use, combine, and extend research performed in the two other axes of SIMILAR (i.e. Axis 1: Theoretical frameworks, and Axis 2: Interaction paradigms) and integrate activities in the following application areas: disabled human-computer interaction, computer-mediated communication of people with disabilities for communication between people with the same disability, technology-enhanced training of people with disabilities (training of the blind and visually impaired, sign language training, dyslexic people training, etc.), and translation between modalities for communication between people with different disabilities.

Speech technology has already opened great opportunities for both the education and social interaction of blind people.

SIMILAR will reinforce this trend, and will address a wider public by enabling the computer to serve as a translation interface between sign language and speech or gestures. This is important in the sense that sign language appears to be essential for communication between the deaf and mute, but also for the intellectual development of young preschool children.

Work within SIMILAR faces various challenging scientific problems and should lead to numerous scientific international publications. Moreover, this WP intends to contribute to ongoing standardisation activities and to produce open source software to foster the development of communication assistants for the disabled.

The European Assistive Technology (AT) market has generally been neglected by industry. The roots of the problem are many: "market fragmentation, technological background, provision needs, non-uniform standards, non-adequate testing facilities, and the social structure of the target group" [1].The integration effort to be made in SIMILAR will support the emergence of a unified and coherent AT market, creating new opportunities for industries.

The chapter is organised as follows: Section 2 briefly presents the multimodal interface principle and also deals with the main concept of work to be integrated within SIMILAR, i.e. the modality replacement principle. The main research areas within SIMILAR for “disability applications” are presented in Section 3, while Section 4 focuses on a “dream” scenario for developments in multimodal interfaces within the next ten years. Finally, conclusions are drawn in Section 5.

6.2. Modality Replacement in Multimodal Interfaces

6.2.1. Multimodal Interfaces

The term “multimodal interfaces” refers to interfaces that also support non-GUI interaction. The multimodal interfaces include many advantages for error recovery, such as:

- Users intuitively pick the mode that is less error-prone.
- Language is often simplified.
- Users intuitively switch modes after an error, so the same problem is not repeated.

For collaborative work, multimodal interfaces can communicate a lot more than text:

- Speech contains prosodic information.
- Gestures communicate emotion.
- Writing has several expressive dimensions.

The challenges in multimodal interfaces mainly refer to the fact that using multimodal input generally requires advanced recognition methods for each mode, for combining redundant information, and for utilising, in the best way, complementary and non-redundant information.

6.2.2. Modality Replacement

The major scientific issue to be explored in SIMILAR is the concept of modality replacement. This is crucial in applications for people with disabilities people who lack one or more of their senses, or people who are unable to use one of their senses (e.g. drivers in cars).

In such cases the use of multimodal information is the only solution. However, many issues still need to be resolved, such as a) the use of early or late fusion of multimodal information, b)

the synchronisation of feature extraction and modality integration procedures, and c) the simplification of the resulting user interfaces in order to be used by disabled users.

The ability of the multimodal systems to substitute for a missing modality is the major requirement of a multimodal interfaces system for the disabled. The system should be able to adapt to the percentage of modality replacement needed for each user and the special needs of his/her disability. Early or late fusion of various modalities should be chosen based on the application, disability, and degree of modality replacement needed.

In order to solve the problem of modality replacement in multimodal interfaces, information is combined at two levels, namely, a) the *feature level* (early fusion), as shown in Figure 19 and b) the *semantic level* (late fusion), as shown in Figure 20.

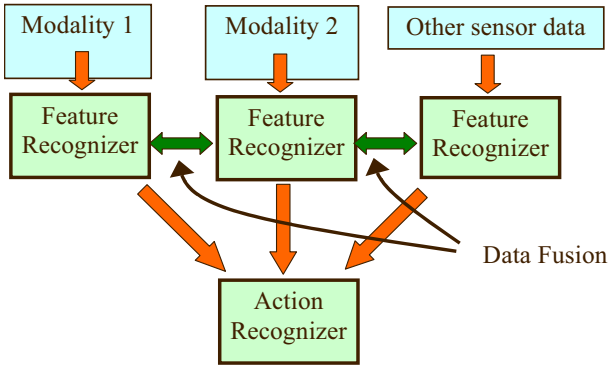


Figure 19 Early fusion of multimodal data

Early fusion applies to combinations such as speech and lip movement. It is considered difficult, in general, mainly because of the need for multimodal training data. Also, the need for data synchronisation and the computational and training costs are important factors that have to be taken into account.

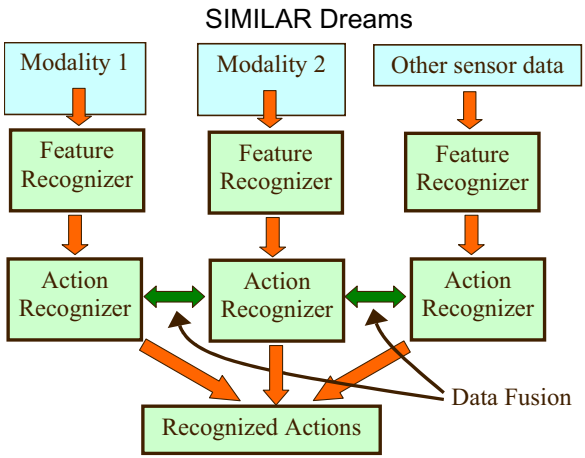


Figure 20. Late fusion of multimodal data

Late fusion is appropriate for combinations of complementary information such as the pen and speech. In this case the recognisers are trained and used separately. Thus, already available off-the-shelf unimodal recognisers can be used in late fusion multimodal interfaces. However, it is still important to accurately time-stamp all inputs: typical delays are known between gesture and speech, for instance.

All aforementioned multimodal interface strategies will be explored within SIMILAR to serve the special need for modality replacement. Data fusion in both the feature and the action level will be used, making the assumption of one missing modality both as input and as output of the procedure.

6.3. Main Research Areas within SIMILAR

The main research areas in multimodal interfaces for people with disabilities within SIMILAR include but are not limited to:

- Applications for the blind: a) A haptic tool to access HTML documents, and b) cane simulation and object recognition applications from virtual environments.

- Applications for the deaf: a) Internet tools for sign language synthesis, b) sign language tutoring, and c) cued speech recognition with a 3D camera.
- Applications for people who cannot move their hands including head pose estimation integrated with eye-blink detection and speech synthesis/recognition for car drivers.

A detailed analysis of the main research foci in the above areas is provided in the sequel.

6.3.1. Haptic Applications for the Blind

Research within SIMILAR will focus on the development of: a) a haptic tool to access presentations and b) virtual interactive environments for cane simulation and object recognition applications.

6.3.2. Haptic Tool to Access Presentations

The greatest potential benefits from virtual environments can be found in applications concerning areas such as education, training, and communication of general ideas and concepts [9]. The technical trade-offs and limitations of the currently developed virtual reality (VR) systems are related to the visual complexity of a virtual environment and its degree of interactivity [10,11]. Hitherto, several research projects have been conducted to assist visually impaired people in understanding 3D objects, scientific data and mathematical functions by using force feedback devices [12]-[16].

In [2], Landua and Wells presented a talking tactile tablet, which combines tactile input with relevant and immediate audio data. The proposed interface improves speed and ease of learning and reinforces learning through dual modalities (haptics and audio).

In [3], recognition of geometrical and general VRML objects is examined. Furthermore, tests about understanding mathematical surfaces and navigation are examined. The outcomes of these

tests show that blind users are able to handle quite complex objects and environments, too, and that realistic virtual environments in some cases appear easier to handle than more abstract test environments.

Three designs of a scrolling function in a Virtual Environment (VE) are examined in [4]. The user can press the sides of a world-limiting box and the world moves so that objects behind the wall are pressed to move into the limiting box. The user can also press the different arrow keys to move the world in the direction of the key, or the PHANToM™ stylus button, and drag the world in the horizontal plane.

Computer games for blind users have been designed in [5]. A two-finger haptic interface is used in GRAB HAVE (Haptic Audio Visual Environment). A haptic geometric modeller enables people to locate and interact with 3D computer-generated objects through touch and audio. A Wb-based haptic tool that enables access to line graphs, bar charts, and pie charts is introduced in [6].

The focus within SIMILAR will be to develop an *interactive presentation environment for visually impaired or blind people*. This application will be aimed at developing a haptic user interface for studying interactive presentations. The users will be able to move a cursor in the 3D virtual environment, select objects, and receive haptic and audio feedback by using a haptic device such as the PHANToM™.

Each presentation may contain various pages. Each page will be a collection of objects placed inside a bounding box. Objects that complement a page of the presentation are drawn inside this area. In order to fulfil the presentation needs, three different types of object will be defined: *inactive objects*, *active objects* and *links* or *buttons*. Collision detection will be performed between the cursor and objects (of any kind) and between the cursor and the sides of the bounding box. Figure 21 presents the structure of the presentations to be processed by the haptic presentation environment.

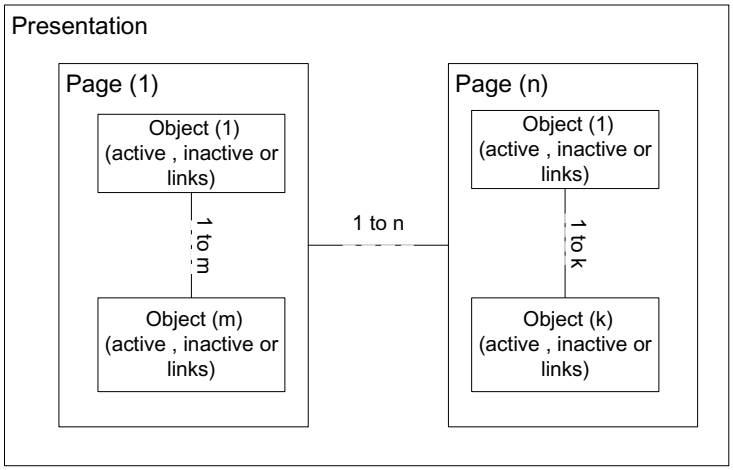


Figure 21. Presentation structure

Active objects will have two main characteristics: i) when the cursor contacts an active object it will declaim a short description of what it represents and ii) when the user presses the stylus button while the 3D cursor resides on the object, it will declaim a detailed description of itself. The links will also have characteristic (i) of active objects. Another characteristic of link objects is that when the user presses the button on the stylus while the cursor resides on a link object, it will declaim a detailed description of itself and a short description of changes that will happen in the presentation environment, and finally change the objects that exist in the current page with new objects necessary to present the selected page.

To allow the user to study the presentation bi-directionally, every page should contain a link object called “back”. The user can return to the previous page of the presentation by clicking on that object. The “back” object will always be placed in the left bottom corner of the bounding box. When the user is on the initial page of a presentation the object is deactivated.

The conversion of a regular presentation to a “haptic accessed presentation” works as follows: An initialisation haptic page is created with one active button that links to the first page of the presentation. The title of the first page is used as the short description of the button and the plain text (without any bullets) of the page as the detailed description. If there is no plain text in the page, the detailed description is the same as the short description. Any links in the plain text are presented as separate link objects in the scene.

If the presentation page contains lists, each element of the list is converted to an active object (Figure 4). The first order elements are converted to active objects and subsequent bullets are connected to these using inactive objects. Each element that contains a link is converted to a link object. Images are converted to active objects and use the alternative image text for the short and detailed descriptions.

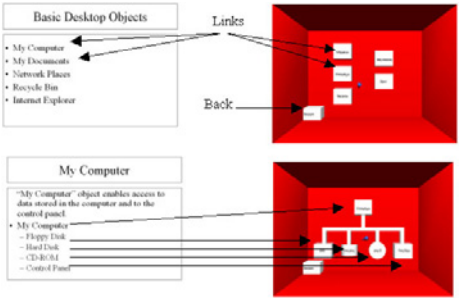


Figure 22. Original HTML presentation pages
and corresponding application pages,
two objects’ shapes in the second case were changed manually.

The conversion stops when there are no more links to convert or there are links to files on the Web.

6.3.3. Virtual Interactive Training Environments for the Blind

6.3.3.1. Cane Simulation

Cane simulation applications will be implemented for realistic navigation tasks with the use of CyberGrasp™, which, in combination with the Ascension MotionStar™ wireless tracker, leads to a significant workspace expansion (up to 7 metres). Cane simulation applications could include indoor and outdoor environments, such as navigating inside a bank or public building, crossing streets (Figure 23), etc.

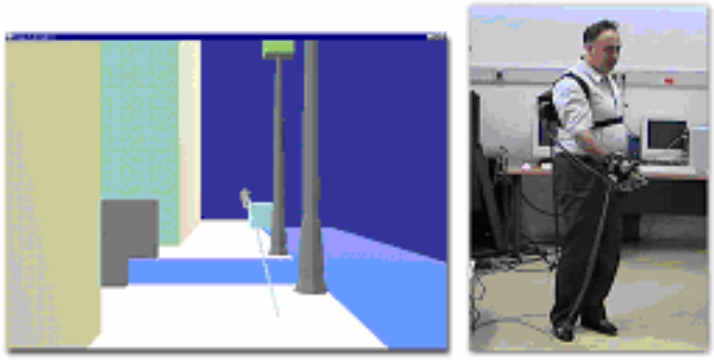


Figure 23. Cane Simulation,
(left: virtual environment,
right: a user).

The cane will be designed to be an “extension” of the user’s index finger. The force feedback applied to the user’s hand will depend on the orientation of the cane relative to the virtual object with which it collides. Specifically, when the cane hits the ground, force feedback will be sent to the user’s index finger. Force feedback will be applied to the thumb when the cane collides with an object on the user’s right and force feedback will be

applied to the middle ring and pinky finger simultaneously when the cane collides with an object on the user's left.

Forces applied to the user can be summarised as a constant continuous force that emulates the force provided by grasping a real cane, a cosine force effect (buzzing) applied to the user when the cane penetrates an object, and a jolt force effect that is sent to the user when the cane hits an object or the ground.

Cane simulation is expected to be a pioneering application that will give rise to new training applications for the visually impaired, especially blind children.

6.3.3.2. Object Recognition

In recent years there has been growing interest in developing force feedback interfaces that allow blind and visually impaired users to access not only two-dimensional graphic information, but also information presented in 3D virtual reality environments (VEs) [7]. It is anticipated that the latter will be the most widely accepted, natural form of information interchange in the near future [8].

Haptic virtual reality can be used to develop training applications for object recognition in which the user can touch/feel the shape of the objects using his/her hand [17]. Two haptic devices are going to be used within SIMILAR, i.e., the PHANToM™, which provides feedback to the index finger, and the CyberGrasp, which provides feedback to the four remaining fingers (Figure 24).

Figure 24: training blind and visually impaired using haptic virtual reality techniques.

The challenge within SIMILAR is to develop novel object recognition applications focusing on providing access to “real” 3D objects. The aim is to use an automatic depth map extraction tool to create 3D models in real time and to use these models in the recognition environment. The purpose is to create a tool that enables blind users to access information in real time.

Initially, the scene will be captured using a monoscopic or stereoscopic camera. In the following, the video is processed and information about scene structure and characteristics is extracted. Using this information a 3D model of the scene is generated and used in order to create a haptic representation of the observed scene. This 3D model of the scene is a pure transformation of the 3D data obtained in the structure reconstruction step. If knowledge about the observed scene is available, the 3D model generation can be extremely accurate. Parameters of the model such as global and local scaling, translation and rotation are estimated directly from the 3D data.

For scenes that include geometrical surfaces of moderate complexity, a superquadric approximation can also be used. This method is widely used in the modelling of 3D data using range

images (depth maps) [18]. A non-linear least squares minimisation method will be used to estimate the parameters of the superquadrics.

6.3.4. Sign Language Analysis

Sign language recognition is a multidisciplinary research area involving pattern recognition, computer vision, natural language processing, and psychology. Sign language recognition is a comprehensive problem not only because of the complexity of the visual analysis of hand gestures, but because of the highly structured nature of sign languages as well.

Although sign languages are well-structured languages with a phonology, morphology, syntax and grammar, they are different from spoken languages. The structure of spoken language makes use of words linearly, i.e., one after the other, whereas sign language makes use of several body movements in parallel in a spatial and temporal space. The linguistic characteristics of sign language differ from those of spoken languages due to the existence of several components affecting the context, such as the use of facial expressions and head movements in addition to hand movements.

A very brief look at sign language grammar illustrates the challenges faced: Sign language phonology makes use of formational parameters such as hand shape, joint position, and movement. The morphology uses directionality, aspect, and numeral incorporation, while the syntax uses spatial aspects such as localisation and spatial agreement as well as facial expressions. It is clear that sign language recognition is a very complex task, one that uses hand shape recognition, gesture

recognition, face and body part detection, and facial expression recognition as its basic building blocks (Figure 25).

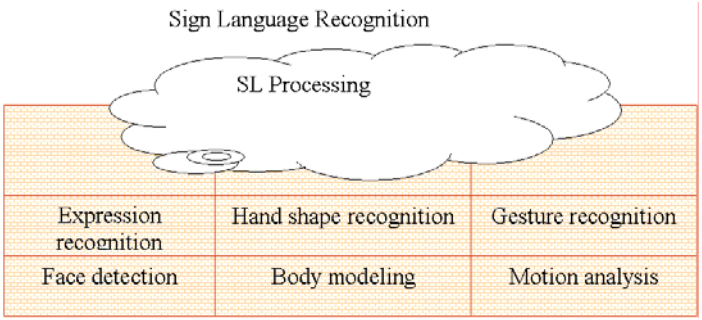


Figure 25. The building blocks of Sign Language Recognition

Sign Language recognition requires both the hand trajectory and hand posture (position, orientation, joint angles) information. To solve the hand trajectory recognition problem, Hidden Markov Models have been used extensively for the last decade. Lee and Kim [19] propose a method for online gesture spotting using HMMs. Starner et al. [20] used HMMs for continuous American Sign Language recognition.

The vocabulary contains 40 signs and the sentence structure to recognise was constrained to personal pronoun, verb, noun, and adjective. In 1997, Vogler and Metaxas [21] proposed a system for both isolated and continuous ASL recognition sentences with a 53-sign vocabulary. In a later study [22] the same authors attacked the scalability problem and proposed a method for the parallel modelling of the phonemes within an HMM framework.

In [23], a stereo system has been used to perform 3D gesture recognition. This work focuses on model-based 3D hand shape recognition and incorporating the shape information into the

gesture recognition system using an input-output HMM. This model-based hand posture recognition system extracts hand shape information as a vector of finger joint angles in the continuous domain.

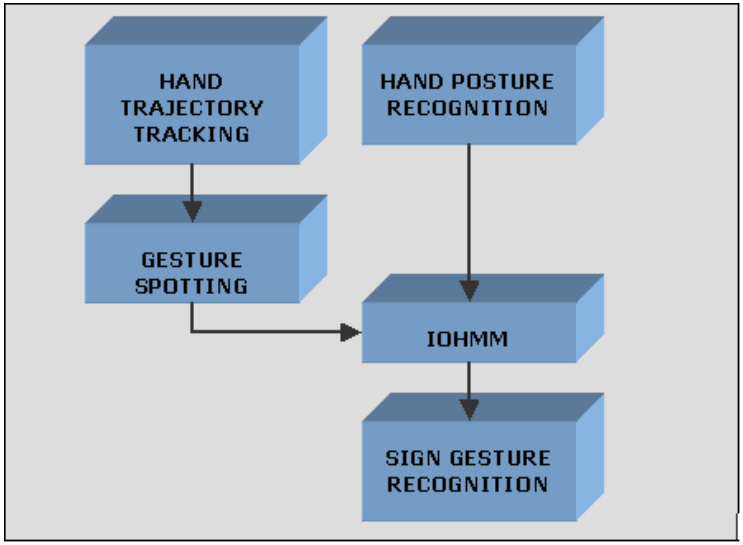


Figure 26. Block diagram of the sign language recognition system

Figure 26 shows the block diagram of the system. Work within SIMILAR will focus on employing this system in an interactive sign language teaching tool. Obviously, for mere recognition purposes such detailed posture information seems unnecessary. However, individual joint angles are needed instead of discrete hand states (such as open or closed states), since the aim is also to reconstruct the gestures performed by the user so as to be able to give feedback in the interactive sign language tutoring tool.

6.3.5. Sign Language Synthesis

A new Web tool for converting text to sign language notation and corresponding VRML animation sequences for H-anim compliant avatars will be designed and implemented within SIMILAR. The tool will be used as both a sign language dictionary and a text for the sign language translator. It will be based on a sign language dictionary written in the well-known SignWriting system. Specifically, the dictionary entries will be in SWML (SignWriting Markup Language), an XML-based format that was recently developed for the storage, indexing and processing of SignWriting notation.

Any SWML dictionary, including those obtained from existing SignWriting dictionaries (e.g., using the online conversion tool available at the SWML site), can be easily supported. For sign synthesis, each sign box (basic sign) will first be converted to a sequence of Body Animation Parameters (BAPs) of the MPEG-4 standard corresponding to the represented gesture. These sequences, which can also be coded and/or reproduced by MPEG-4 BAP players, will then be used to animate H-anim compliant VRML avatars, reproducing the exact gestures in sign language notation.

Envisaged applications include the development of signing avatars for interactive information systems (Web, E-mail, info–kiosks) and TV newscasts for persons with hearing disabilities. Figure 9 shows an example of a user interface for the sign language synthesis tool.

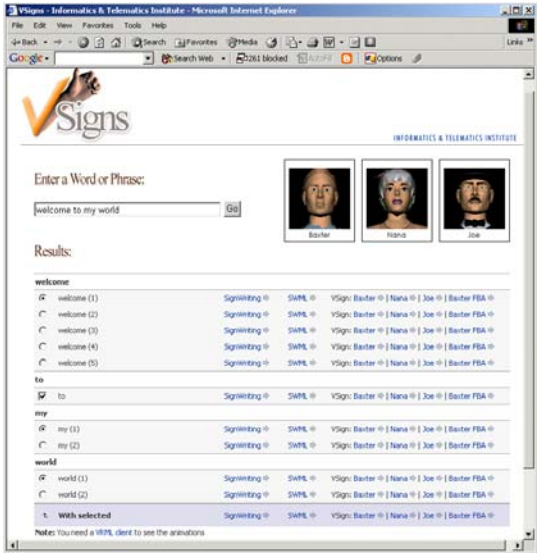


Figure 27: Results of “Welcome to my world” query.
The user is able to select the desired terms and then produce and display sign synthesis results using the selected words or the entire phrase, using any of the available H-anim avatars.

An ambitious extension of the system to support non-manual body movements as well as facial animation is also under development. Facial animation will be represented by MPEG-4 Facial Animation Parameters, which will be reproduced using H-anim compliant avatars performing simultaneous face and body animation. A problem with using Facial Animation Parameters is that most of them, in contrast to BAPs, describe complex non-rigid motions, and therefore most existing FAP player implementations are model-dependent. Furthermore, the resulting VRML animations are more complicated since they contain numerous CoordinateInterpolator nodes (one per face model vertex). Therefore, the computational demands on the hardware that reproduces these animations are increased.

6.3.6. Cued Speech Recognition

Deaf people can use several languages to communicate with other people. One is the well known sign language and the other is cued speech. Cued speech consists of a combination of lip shape and motion and hand gestures. Lip reading is not always enough to understand speech completely. Hand gestures are then used to remove lip shapes ambiguities. For example, “p” and “b” yield the same lip shape and motion. Those two letters are thus associated with very different hand gestures. Cued speech hand gestures are used in order to code phonemes. Each phoneme is the combination of a hand configuration coding the consonant and a hand position with respect to the face coding the vowel. Cued speech has been defined for many languages (including all European languages) and the gestures are the same, with a few added gestures for the most complex languages. There are eight hand configurations and five hand positions (see Figure 10 for the French version of cued speech).

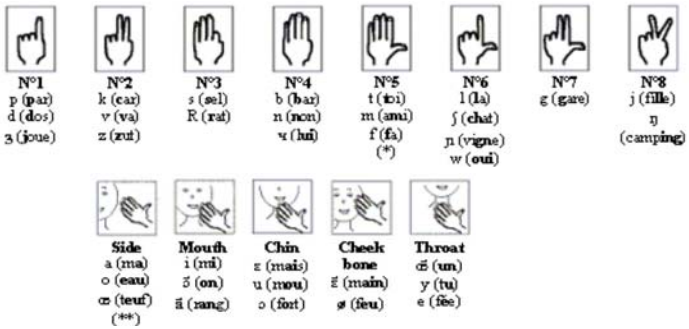


Figure 28: t: the French version of cued speech

Computerised systems can be used to enable or enhance communication between deaf people and hearing people. Two applications are considered: face to face and phone

communication between deaf and hearing persons. In those applications our basic assumptions are that the deaf people are not able to produce speech and that hearing people are not able to understand or produce cued speech.

- Deaf to hearing people communication: The computer system has to recognise lip shape and motion and cued speech hand gestures in order to produce speech
- Hearing to deaf people communication: The computer system has to recognize speech and to synthesise lip shape and motion and cued speech gestures

Most work on cued speech to date has concerned automated cued speech teaching and synthesis [24]-[26]. As far as we know, no work has been published about automatic cued speech analysis and recognition.

In SIMILAR we shall address the following research subjects:

- hand gesture analysis applied to cued-speech hand gestures. Specifically we shall compare classification based on 3D information and 2D segmentation mask information
- joint analysis of hand gesture and lip shape and motion analysis thanks to evidence theory, including dynamic aspects of the gestures
- temporal segmentation of cued-speech hand gestures in order to identify the beginning and end of gestures

6.4. The dream scenario

The following scenario is based on the integration of various multimodal interface technologies for communication between people lacking different modalities.

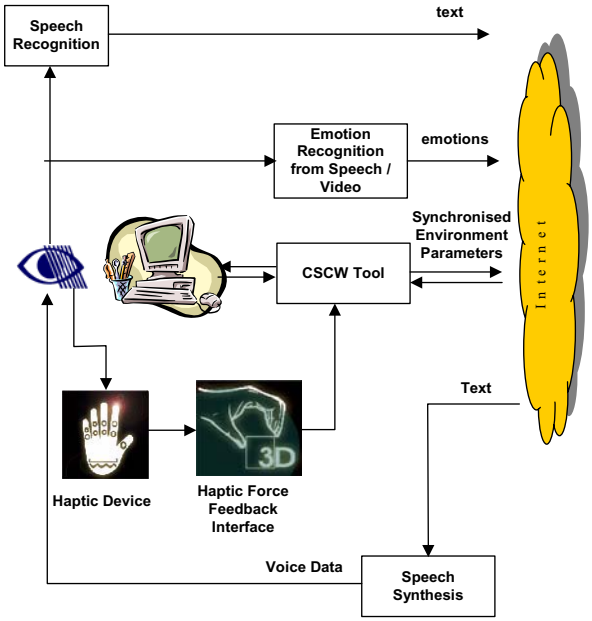


Figure 29: Block diagram of a CSCW tool for the collaboration of remotely located people with disabilities (Part 1 – Blind person side).

Figures 29 and 30 show the dream scenario for the next ten years in multimodal interfaces for the disabled. Specifically, the scenario presents a blind and a deaf user collaborating remotely using a Computer Supporting Cooperative Work (CSCW) tool.

The blind user provides input to the tool via speech, haptics, and emotions, while he/she receives input from speech. The deaf user provides input to the tool via sign language and receives information through sign language and emotion synthesis via an avatar. Speech recognition and synthesis, emotion recognition from speech and video, text-to-sign language synthesis, emotion synthesis, avatar animation, and sign language analysis technologies are integrated in the proposed scenario.

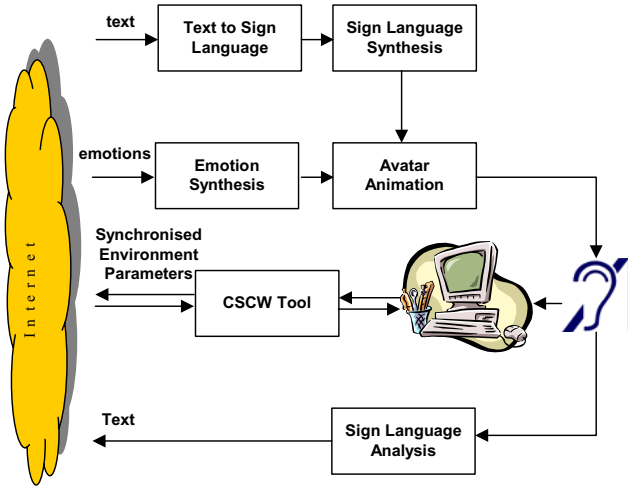


Figure 30 Block diagram of a CSCW tool for the collaboration of remotely located people with disabilities (Part 1 – Deaf person side).

6.5. Conclusions

This chapter presented in detail the grand challenges in multimodal interfaces research in applications for people with disabilities. The basic development concept in SIMILAR is the idea of modality replacement, i.e., the use of information originating from various modalities to compensate for the missing input modality. A number of new applications have been presented in brief, such as the use of haptics and speech for the visually impaired in applications such as computer access and cane simulation for training, sign language analysis and synthesis, and cued speech recognition. Finally, some “dream” applications have been presented, such as collaboration between remotely located people with disabilities exhibiting various types of disabilities.

6.6. References

Carlos A. Velasco, “The Information Society disAbilities Challenge (ISdAC): Paving the way for the active end-user”, www.stakes.fi/tidecong/844isdac.html.

Landua, S. and Wells, L., “Merging Tactile Sensory Input and Audio Data by Means of The Talking Tactile Tablet”, Proc. Eurohaptics 2003. Dublin, Ireland, (2003)

Magnusson, C., Rassmus-Gröhn, K., Sjöström, C., Danielsson H, “Navigation and recognition in complex haptic virtual environments – reports from an extensive study with blind users”, Proc. Eurohaptics 2002, Edinburgh, UK, (2002).

Magnuson, C and Rassmus-Gröhn, K., “Non-visual Zoom and Scrolling Operations in a Virtual Haptic Environment”, Proc. Eurohaptics 2003, Dublin, Ireland, (2003).

Wood, J., Magennis, M., Cano Arias, E. F., Gutierrez, T., Graupp H. and Bergamasco, M., “The Design and Evaluation of a Computer Game for the Blind in the GRAB Haptic Audio Virtual Environment”, Eurohaptics 2003, Dublin, Ireland, 2003.

Yu, W., Cheung, K., Brewster, S., “Automatic Online Haptic Graph Construction”, Eurohaptics 2002 Edinburgh, UK, 2002.

Colwell C., Petrie H., Kornbrot D., Hardwick A. and Furner S., “Haptic Virtual Reality for Blind Computer Users”, in Proc. of Annual ACM Conference on Assistive Technologies (ASSETS '98), pp 92-99, (1998).

Burdea G. C., “Force and Touch Feedback for Virtual Reality”, Wiley-Interscience Publication, (1994).

C. Sjostrom, “Touch Access for People With Disabilities”, Licentiate Thesis, in CERTEC Lund University, Sweden, (1999).

C. Sjostrom, “Designing Haptic Computer Interfaces for Blind People”, in Proc. of ISSPA 2001, Kuala Lumpur, Malaysia, August (2001).

C. Sjostrom, "Using Haptics in Computer Interfaces for Blind People", in Proc. of CHI 2001, Seattle, USA, March (2001).

V. Scoy, I. Kawai, S. Darrah, F. Rash, "Haptic Display of Mathematical Functions for Teaching Mathematics to Students with Vision Disabilities", Haptic Human-Computer Interaction Workshop, (2000).

P. Penn, H. Petrie, C. Colwell, D. Kornbrot, S. Furner and A. Hardwick, "The Perception of Texture, Object Size and Angularity by Touch in Virtual Environments with two Haptic Devices", Haptic Human-Computer Interaction Workshop, (2000).

J. P. Fritz and K. E. Barner, "Design of a Haptic Data Visualisation System for People with Visual Impairments", IEEE Transactions on Rehabilitation Engineering, vol. 7, no. 3, pp. 372-384, (1999).

N. A. Grabowski and K. E. Barner, "Data Visualisation Methods for the Blind Using Force Feedback and Sonification", SPIE-Int. Soc. Opt. Eng. Proceedings of Spie - the International Society for Optical Engineering. vol. 3524, pp. 131-139, (1998).

F. Van Scoy, T. Kawai, M. Darrah and C. Rash, "Haptic Display of Mathematical Functions for Teaching Mathematics to Students with Vision Disabilities: Design and Proof of Concept", Haptic Human-Computer Interaction. Springer LNCS, Vol 2058. pp 31-40, (2000).

D. Tzovaras, G. Nikolakis, G. Fergadis, S. Malasiotis and M. Stavrakis, ``Design and Implementation of Haptic Virtual Environments for the Training of Visually Impaired'', *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, Vol. 12, No. 2, pp.266-278, June 2004.

F. Solina and R. Bajcsy, "Recovery of Parametric Models from Range Images: the Case for Superquadrics with Global Deformations", IEEE Trans. Pattern Anal. Mach. Intell, Vol. 12, No. 2, pp.131-147, 1990

H. Lee and J.H Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 961-973, 1999.

T. Starner, J. Weaver, and A. Pentland. Real-time American Sign Language recognition using desk and wearable computer based video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(12):1371–1375, December 1998.

C. Vogler and D. Metaxas. Adapting hidden markov models for ASL recognition by using three-dimensional computer vision methods. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, pages 156–161, Orlando (USA), 1997. IEEE Computer Society Press.

C. Vogler and D. Metaxas. Parallel Hidden Markov Models for ASL Recognition, Proceedings of the International Conference on Computer Vision, Krkyra, Greece September 22-25,1999

C. Keskin, A.N. Erkan, L. Akarun, Real time hand tracking and 3D gesture recognition for interactive interfaces using HMM, Proceedings ICANN/ICONIP, Istanbul, 2003.

Development of speech reading supplements based on automatic speech recognition -Duchnow ki P., Lum D.S., Krause J.C., Sexton M.G., Bratakos M.S., Braida L.D.- IEEE Transactions on Biomedical Engineering, April 2000,47(4):487-496.

A multi-media program exercising the basics in lip-reading, cued-speech and sign- language vocabulary – Cook B..- Speech and Language Technology for Disabled Persons. Proceedings of an ESCA Workshop.1993:pp.79-82.ESCA “Speech Commun “Music Acoust. KTH, Stockholm, Sweden.

French lip-reading and cued-speech training by interactive video – Vandenbemden G., Dufour P., Marco C. Computers for Handicapped Persons.1989: pp. 359-366.Vienna,Austria.

Chapter 7: Migratory Multimodal UI

Fabio Paternò (ISTI, Italy)

Nowadays, many technological innovations are continuously appearing on the market, making modern life a multi-platform environment in which people are surrounded by different types of devices through which they can connect to networks in different ways and can interact in different modalities while moving about freely through different contexts.

Multimodal user interfaces play an important part in enabling users to exploit such possibilities fully. However, the design of such multimodal user interfaces must take into account the features of the interaction resources of the devices currently available.

Migratory interfaces are interfaces that allow users to change devices freely and still be able to continue the task from the point they left off in the previous device. In order to obtain a satisfying interaction three aspects are important:

- Interaction Continuity: Continuity at the task performance level should be guaranteed.
- Platform Adaptation: The user interface should adapt to the new device's features.
- Interface Usability: The user interface should apply design criteria for obtaining usable results.

The combined use of migration and multimodal interfaces is fundamental to obtain environments able to support natural interaction, where users can move about freely and still continue their activities in an effective and efficient manner. These features can be important in many types of applications: financial, on-line auctions, games, and whenever users have to change place but would still like to carry on their tasks without having to start from scratch.

Figure 31

7.1. Process

Figure 32 shows the steps in the migration process, where the ovals are the activities and the rectangles are the data manipulated and produced in the different phases.

One key aspect in the migration process is the ability to capture the state of the migrating interface, which is the result of the history of user interactions with the application, including visited

pages, submitted data, and the results of previous data processing.

We have developed a migration engine that collects information on the context, user interface, interactive system, and migration requests.

At the beginning, the migration engine uses all this information to calculate the migration mapping and to store the interface state. These two processes produce the interface state, which is composed of all the user interface run time data, and identification of the devices involved, with the associated interaction resources and related interaction objects to be used. Then, the migration engine adapts the user interface and its state to the target devices following user preferences. The result of this process is the specification of the user interface adaptation to perform. Finally the run-time support generates the corresponding user interface.

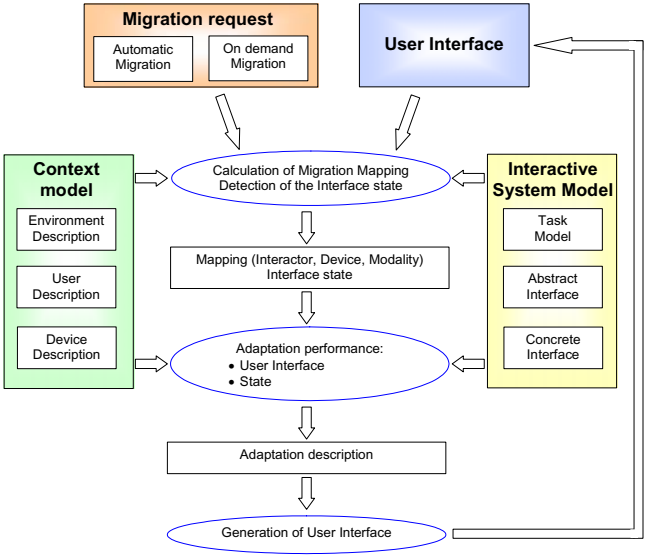


Figure 32

7.2. Architecture

In our migration environment, first of all the devices to be involved in any migration have to register with the migration service. In this way, the migration server has a list of available devices and additional information regarding their features (platform, availability, whether it is suitable for shared use, etc.). The migration requests can be issued by either the user or the system (for example, when it detects that the battery is low).

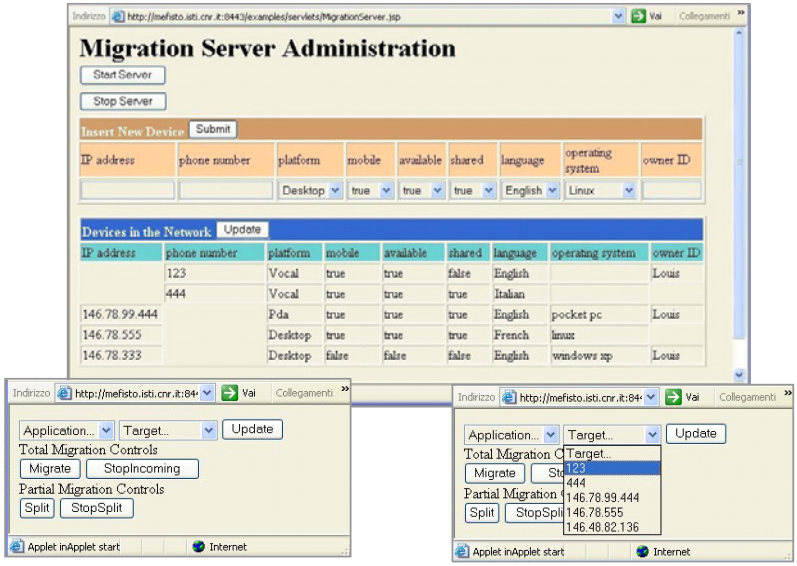


Figure 33

In order to allow users to send migration requests, a client application with its user interface should be loaded and running. When migration is requested, the environment is able to collect the state of the source user interface resulting from the previous user interactions (such as text entered, elements selected, and so on).

Then, the user interface is dynamically adapted to the identified target device's features. This version is transformed in such a way as to maintain the state resulting from the previous interactions and adapt it to the features of the target interface as well.

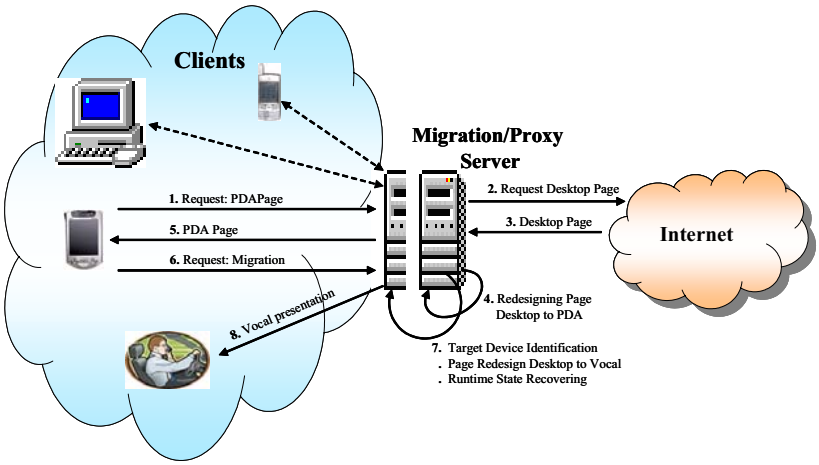


Figure 34

The latest version of our environment is based on a migration/proxy server that supports migratory interfaces in multi-device environments. Starting with a pre-existing version for desktop platforms, the environment is able to generate interfaces for different platforms and modalities dynamically, exploiting the semantic information that is derived through reverse engineering techniques. The environment currently supports access to Web applications and is able to generate versions for PDAs, various types of mobile phone, vocal devices, and multimodal devices, and then supports migration through them.

7.3. HIIS profile

The "Human Interfaces in Information Systems" Laboratory (HIIS) of the Information Science and Technologies Institute (ISTI) is an institute of the Italian National Research Council (CNR).

Our main goal is to carry out research in methods and tools to support user interface designers and software developers to obtain usable systems that can be accessed from different contexts of use. More specifically, our research interests focus on:

- Design of Interactive Systems
  - Model-based Design
  - Tools for Task Modelling and Analysis
  - End-user Development
- Usability and Accessibility
  - Remote Usability Evaluation
  - Usable and Accessible Web Sites
- Mobile Information Systems for All
  - Multiplatform multimodal interface generation
  - Design of Location-aware, Indoor PDA Guides

Web site: www.isti.cnr.it/ResearchUnits/Labs/hiis-lab/

SIMILAR Dreams

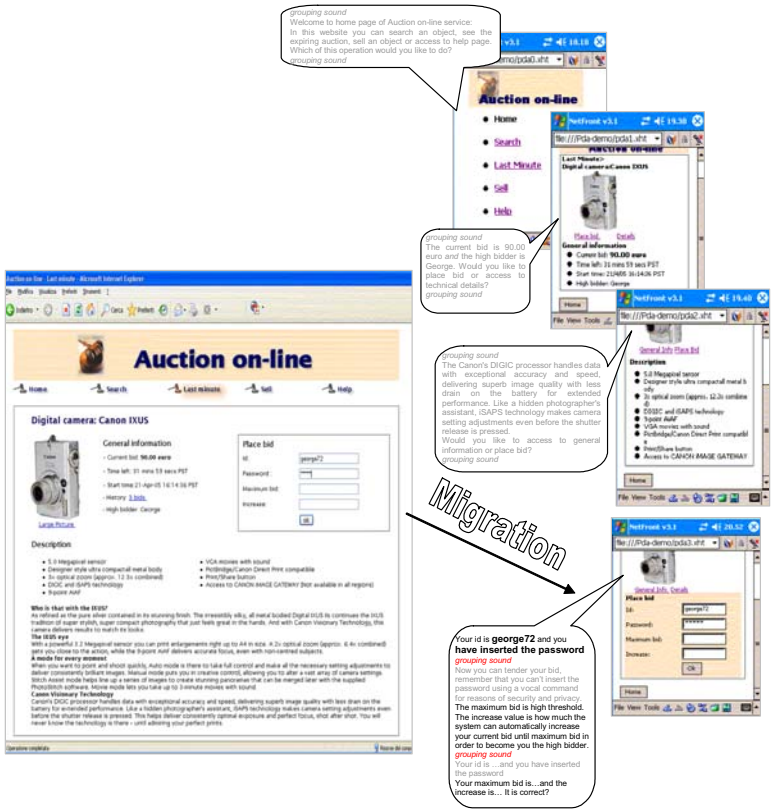


Figure 35

Part 2: Achievements

Chapter 8: SIMILAR work ’04

Benoit Michel (UCL, Belgium)

The young SIMILAR network of excellence started in 2004 with integration and dissemination activities as well as a number of research tasks. R&D tasks were segmented in several work-packages corresponding to the various levels of the SIMILAR view on multimodal interface research, as we can see in the diagram below:

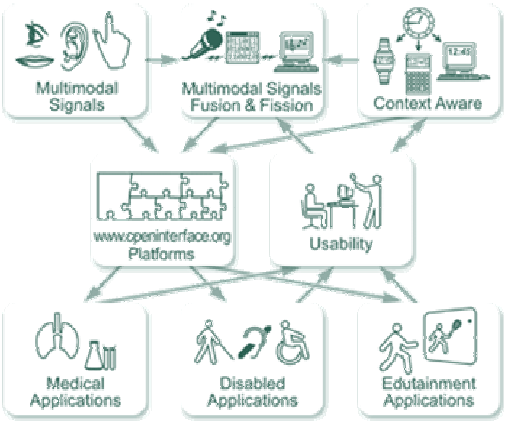


Figure 36

1.1 Integration through eNTERFACE

SIMILAR is preparing the launch of the eNTERFACE'05 summer workshop in Mons, Belgium. This workshop will gather a large group of multimodal interface developers for three weeks on the FPMS campus in Belgium in July 2005. A web site has been set up where all details and the schedule: www.enterface.net .

The call for projects was launched at the end of September '04, and closed in March '05.

Nine projects proposals were received:

1. Combined Gesture-Speech Analysis and Synthesis (Koc University, Istanbul). This project includes the preparation of a database, study of correlations between speech features and gesture units, modelling of gesture units, and adaptation of units for specific speakers.

2. Multimodal Caricatural Mirror (UCL, Louvain-la-Neuve). This project aims at creating a caricatural mirror where people can see their own emotions amplified (image + speech) by an avatar on a wide screen facing them. It includes multimodal face tracking, multimodal emotion recognition, and multimodal emotion synthesis.

3. Biologically driven musical instrument (UCL, Louvain-la-Neuve). A virtual musical instrument driven by EMGs, EEGs, heartbeats, video, etc.

4. Multimodal Focus Attention Detection in an Augmented Driver Simulator (UCL, Louvain-la-Neuve). This project proposes to use physiological signals, the analysis of facial expression, and tracking of the user's focus (eye tracking) in an augmented reality driver simulator.

5. Multimodal learning assistance (INPG, Grenoble). The aim of the project is to develop a multimodal system for the analysis and interpretation of any e-learning user. After inferring the (application-dependent) mental state of the student from his/her speech/gestures/ head movements/facial expressions, the application will try to react with the appropriate speech and mind attitude

6. Exploratory bibliographical project on multimodal interfaces (France Telecom RTD) to bring together pieces of information from different sources, especially from the 1) ergonomics and human-machine interfaces and 2) mobile software engineering communities; and to generate new ideas for research related to these new subject

7. Creation of a multimodal (face, lips, speech, writing, body movement) database for robust speaker identification/verification (University of Crete). This project will take advantage of the fact that SIMILAR partners from several EU countries (hence, with various mother tongues) will attend eNTERFACE.

8. On site tests of the various scenarios of fusion between various modalities and under various noise conditions (University of Crete). The project will be concluded by suggestions on the next steps in fusion R&D.

9. Speech Conductor (LIMSI-CNRS Paris). The Speech Conductor project aims at developing a gesture interface for driving ("conducting") a text-to-speech synthesis system. Then, automatic speech synthesis will be modified in real time according to the gestures of a "Speech Conductor". The Speech Conductor will add expression and emotion to the speech flow using speech signal modification algorithms and gesture interpretation algorithms.

Seven of these proposals were accepted. As far as participation is concerned, the call, the duration of which has been extended, has currently allowed the selection of most of the 40 researchers who will take part in the eNTERFACE '05 summer workshop.

Approximately 20% of these selected participants are women.

1.2 Integration through exchanges and twinnings

The SIMILAR post-doc fair has two facets: a physical meeting held each year between partners lead scientists and continuous information exchange in a dedicated part of the SIMILARnet virtual private network dubbed "Researchers' Mobility".

The 1st post-doc fair has been organised at Louvain-la-Neuve on January 26-28, 2004. The different SIMILAR partners presented their activities in details:

- list of PhD students involved in each activity in each group,
- list of post-doctoral fellows,
- list of PhD students and post-docs ready for short-term exchanges in other groups within SIMILAR,
- open positions in the different groups.

This event allowed a better understanding of who is who in the network, and a first step towards constructive, fruitful exchanges of personnel for short-term stays as well as for permanent positions.

A new edition of the SIMILAR post-doc fair will take place, probably in June 2005.

The PhD twinning is a new mechanism set up by SIMILAR that will initiate pairs of PhD theses “twinned” in a cross-disciplinary way (example: two PhDs on a single multimodal topic with visual and speech aspects). Each of the doctoral students travels to his/her twin’s institution for a short- to medium-term stay and thus both of them will work together on the multimodal sides of their joint work for a few months during their theses. This mechanism leaves each PhD student enough flexibility and individuality in his/her own work to let him/her prove his/her excellence while benefiting from the cross-disciplinary scope enlargement provided by the twinning.

This mechanism has been very successful and is one of the major achievements of SIMILAR so far. Indeed eighteen stays of PhD students in a partner institution were initiated in 2004.

1.3 Public awareness through SIMILAR website

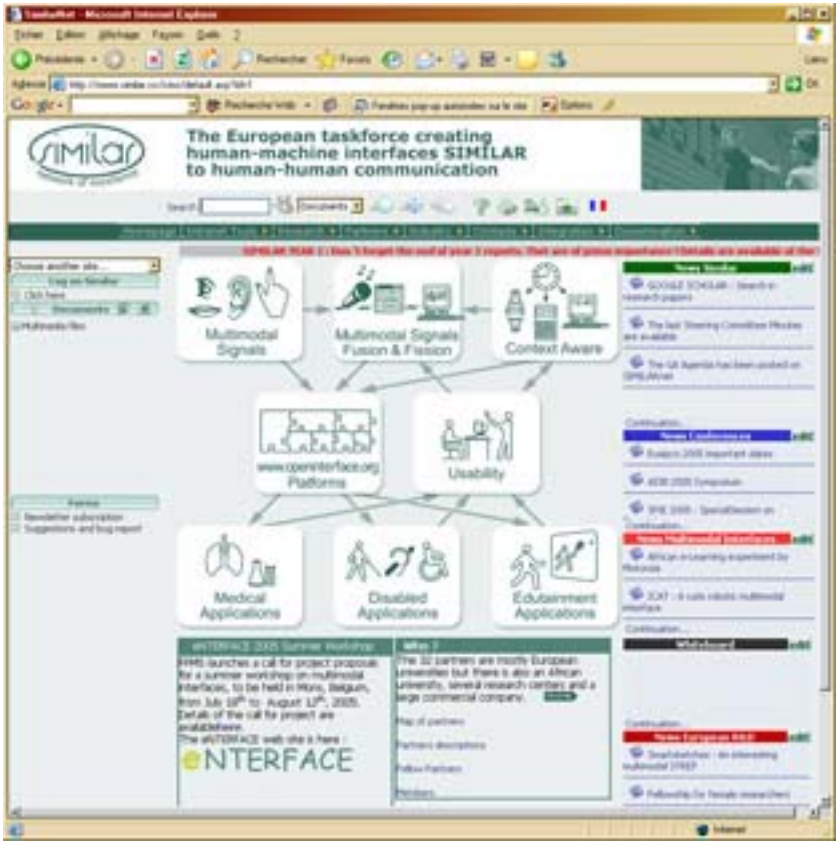


Figure 37: SIMILAR homepage

The SIMILARnet web portal currently has more than 300 subscribed researchers, with a rate of female researchers approaching 30% (71 women for 237 men).

It gets from 870 to 2,401 monthly visits, for a total of more than 20,000 visits for the past year.

Over 350 documents are shared through this website.

The documents and news are well consulted, The site has over 190 documents and more than 100 news.

Each SIMILAR member continues to strive to maintain and improve these figures by keeping the website up-to-date with the latest resources and information on multimodal interfaces.

1.4 Public awareness through the SIMILAR newsletter

SIMILAR has now acquired a new mean of communication, the SIMILAR newsletter.

This publication is the first result of our having set up a SIMILAR editorial board, which is a four-person (three women and one man) team tasked with promoting the image of SIMILAR inside and outside the network (editors@similar.cc).

The SIMILAR newsletter is dedicated to encouraging better mutual knowledge and consequently better communication among SIMILAR members through presentation of the labs and their activities. But it also aims to enhance the visibility of SIMILAR activities to the outside world in order to foster other collaborative undertakings, with a special focus on industry.

The newsletter exists in two versions: paper and electronic.

The paper version is a postcard format that presents the main subjects developed in the Newsletter edition.

The Web version is published on the SIMILAR e-Portal (www.similar.cc). It contains the links to all articles and Internet sites related to of the various headings of the Newsletter.

Three editions of the SIMILAR newsletter have been published to date, the first one in December 2004, the second one in April 2005 and one in June 2005.

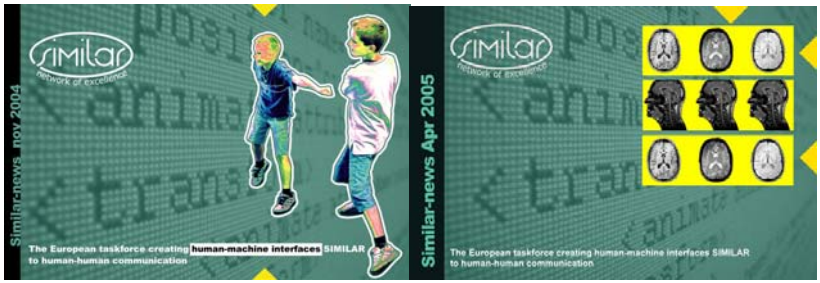


Figure 38: SIMILAR newsletters 1 & 2

The editorial team's goal is to put out a newsletter each two months.

Besides being online and freely available on the SIMILAR e-portal, the newsletter is sent:

- to each SIMILAR member,
- to all members of the Industrial Board
- and to the state organisations (CEC).

Currently roughly one thousand copies of SIMILAR newsletter(paper version) are printed.

1.5 Public awareness through conferences

Due to the overwhelming amount of already existing events, SIMILAR will not organise specific conferences but will play an important role in many EURASIP, IFIP, and IEEE key conferences.

MLMI05 (http://groups.inf.ed.ac.uk/mlmi05/), the 2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms, which will take place on 11-13 July 2005, Royal College of Physicians, Edinburgh, UK, is supported by SIMILAR.

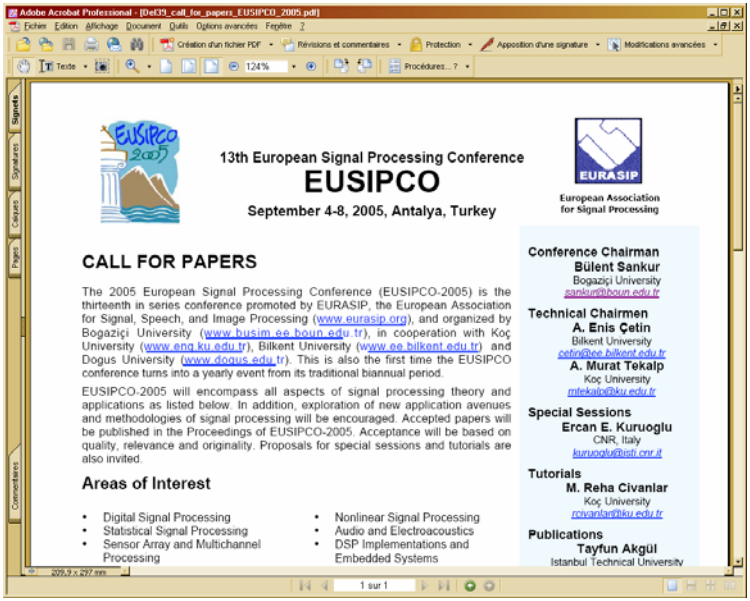
SIMILAR, through its members Benoit Macq and Laurence Nigay, also participates in the MLMI programme committee.

The SIMILARnet intranet will be the communication tool of choice inside the multimodal research community to gather information and prepare the schedule of future participation in conferences, to negotiate workshop acceptance at IEEE and other conferences, to call for papers and submit them, etc.

Our first large-scale participation was at **SPECOM 2004** in Saint-Petersburg, Russia in September 2004. SIMILAR was on the Programme Committee (B. Macq, A. Gilloire, F. Marques, T. Dutoit, and B. Michel) and the network organised two special sessions on multimodal interfaces, speech and natural language, multimodal services and applications for people with disabilities. More information can be found at www.spiiras.nw.ru/speech.

Figure 39: SPECOM 2004

Our second large target is the **EUSIPCO** conference. It was too late to start activities for Eusipco 2004. However, talks about EUSIPCO 2005 were held during the conference and it was decided to have a special session on multimodal interfaces organized by SIMILAR in Antalya, Turkey on September 5th, 2005. Our contacts will be Ferran Marques (UPC) and Bulent Sankur (BUMM). The call for papers was published in December 2004.



13th European Signal Processing Conference

EUSIPCO

September 4-8, 2005, Antalya, Turkey

European Association for Signal Processing

CALL FOR PAPERS

The 2005 European Signal Processing Conference (EUSIPCO-2005) is the thirteenth in series conference promoted by EURASIP, the European Association for Signal, Speech, and Image Processing (www.eurasip.org), and organized by Bogaziçi University (www.busim.ee.boun.edu.tr), in cooperation with Koç University (www.eng.ku.edu.tr), Bilkent University (www.ee.bilkent.edu.tr) and Dogus University (www.dogus.edu.tr). This is also the first time the EUSIPCO conference turns into a yearly event from its traditional biannual period.

EUSIPCO-2005 will encompass all aspects of signal processing theory and applications as listed below. In addition, exploration of new application avenues and methodologies of signal processing will be encouraged. Accepted papers will be published in the Proceedings of EUSIPCO-2005. Acceptance will be based on quality, relevance and originality. Proposals for special sessions and tutorials are also invited.

Areas of Interest

- Digital Signal Processing
- Statistical Signal Processing
- Sensor Array and Multichannel Processing
- Nonlinear Signal Processing
- Audio and Electroacoustics
- DSP Implementations and Embedded Systems

Conference Chairman
Bülent Sankur
Bogaziçi University
sankur@boun.edu.tr

Technical Chairmen
A. Enis Çetin
Bilkent University
cetin@ee.bilkent.edu.tr
A. Murat Tekalp
Koç University
mtekalp@ku.edu.tr

Special Sessions
Ercan E. Kuruoglu
CNR, Italy
kuruoglu@isti.cnr.it

Tutorials
M. Reha Civanlar
Koç University
rcivanlar@ku.edu.tr

Publications
Tayfun Akgül
Istanbul Technical University

Figure 40: Call for papers EUSIPCO 2005

SIMILAR members participated in several other conferences in 2004: Our WP8 leader Prof. J. Vanderdonckt was general conference chairman of both the **ACM Conference on Intelligent User Interfaces IUI'2004** and the **4th International Conference on Computer-Aided Design of User Interfaces CADUI'2004** . More than half-a-dozen other high ranking conferences were on the list.

The title of “best conference paper” presented during the **4th International Conference on Web Engineering ICWE’04** (Munich, 28-30 July 2004) was awarded to J. Vanderdonckt, A. Beirekdar and M. Noirhomme-Fraiture for their “Automated Evaluation of Web Usability and Accessibility by Guideline Review”

1.6 Public awareness through books and papers

Books

A large number of publications related to multimodal interfaces research were published in the first six months of SIMILAR. The two book collections are already under way, with four books published so far. The SIMILAR collection books have official ISBN numbers and may be downloaded for free in PDF format. One can download them or buy them on Internet on the i6doc web site (www.i6doc.com) for a very democratic price (typically under 10 euros, depending on the number of pages).

Figure 41: SIMILAR books

Papers on multimodal interfaces research

The publication list is now 184 papers strong, not counting the two technical papers for internal dissemination only (but with three to be published in 2005).

Figure 42: Some SIMILAR publications

1.7 Public awareness through BEST Summer School

Figure 43: Logo of Louvain-la-Neuve BEST local group
& picture of BEST Summer School 2004 in Louvain-la-Neuve

BEST, the 'Board of European Students of Technology' (website: www.best.eu.org/), is an organisation managed by students that strives to make technology students more internationally minded by encouraging their mobility and intercultural communication. BEST has 1700 active members in 64 local groups (several in SIMILAR partners institutions) and organises 45 summer courses all over Europe.

A BEST summer school was organised from 18 to 31 July 2004 on UCL premises by the local BEST group of Louvain-la-Neuve (www.fsa.ucl.ac.be/BestWWW/FR/index.php?ref=accueil).

SIMILAR partly sponsored it.

The topic of this summer school was 'Medical imaging for diagnosis, image-guided surgery and drug discovery'. There were 18 participants from all over Europe (town of origin: Barcelona, Belgrade, Bucharest, Copenhagen, Ekaterinburg, Krakow, Lisbon, Ljubljana, Naples, Patras, Riga, Skopje, Tallin, Thessaloniki, Valladolid, Warsaw, Zagreb), among whom10 female students.

This was thus perfectly in the scope of SIMILAR, with a focus on the issue of attracting more women to scientific research into multimodality.

The summer school began with an overview of SIMILAR, and most of the presentations were given by SIMILAR researchers.

SIMILAR is willing to reiterate such successful event sponsoring each time the opportunity crops up.

1.8 Public awareness through a software contest

The 'Loco Mummy' is an acronym for "**lo**w c**o**st **mu**lti**m**odal **i**nterface". The idea is to improve the human-human and human-machine interactions without any additional cost on existing hardware on standard desktop PCs, laptops, and/or PDAs by motivating students to work in these areas by the prospect of

winning an award. SIMILAR has set up a SIG dedicated to this goal (the Loco Mummy SIG) where all researchers with related goals will be able to contribute, but it targets mainly last-year students in HCI and signal processing.

Standard PC interfaces are not currently integrating gesture, handwriting, and facial emotion recognition, and other capabilities that could already be possible with the existing input modalities. On the output side, the use of facial reconstruction, virtual avatars, and voice synthesis will lead to better usability of the communication link while consuming only part of the available bandwidth in current computers. A Standard W3C compliant Internet browser with Open Source Java code will be able to offer a far more usable communication experience by adding only a modest amount of artificial intelligence that will be well within reach of modern CPUs.

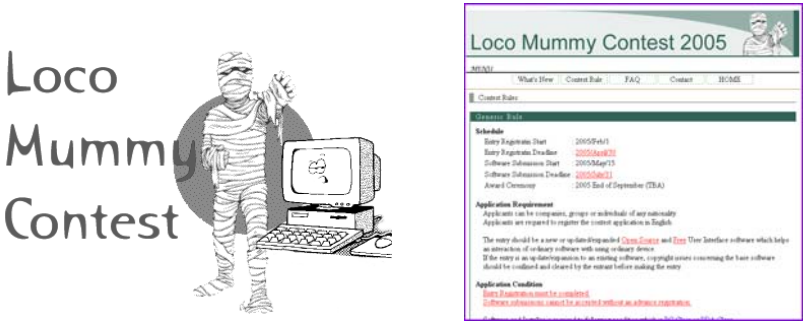


Figure 44: T: the www.locomummy.net logo and web site

The Loco Mummy software contest has its own website and registration for the first event opened on 15 Feb 2005, with a deadline for submissions in October and the award ceremony in December 2005. The event could become recurrent and is seen by many as a good way to promote multimodal interface

research and to attract male and female young scientists to this field.

1.9 SIG on information fusion and fission

The goal of the fusion-fission SIG is to develop a consistent theoretical basis for understanding and formalising co-ordination and co-operation between different modalities involved in the same multimodal interface. WP6 is divided into two main tasks: “Multimodal sources representation in a human-computer interaction context” and “Signal processing for inter-modalities co-operation”.

Multimodal sources representation in a HCI context

The main focus of our research is multimodal communications using the apparatus of distributed source coding with side information and channel coding with side information This activity has just started but we have already achieved some promising results that shed more light on the achievable potential of multimodal systems. We hope that the careful and systematic development of an approach based on the information theory will give us more insights into the optimal design of multimodal management and communications systems in a number of applications. Therefore, the first achievements can be summarised as follows:

- theoretical study of state-of-the-art developments in multi-user information theory as the relevant theoretical basis for multimodal communications;

- investigation of multiple access channels for visual communication within natural scenes, use of side-information, determination of upper-capacity bounds and achievable gains in using multimodal channels, and theoretical evaluation of the informative content of images and the corresponding description complexity;

- development of the first theoretical model of human-to-machine communication protocol based on broadcast channels for speech, lips, gestures, text, and drawing sub-channels; extension of these results to the broadcast channels with side information as part of the analysis of particular codebooks for speech, lips, gestures, text, and drawing sub-channels;

- development of the novel concept of Smart Media, i.e., media that carry hidden additional information for multimedia management and security, and extensions to smart room and electronic- and mobile-commerce applications;

- extension of recent advances in stochastic image modelling and the information-theoretic concept of distributed source coding to multimedia storage, compression, and transmission in unimodal and multimodal systems.

Distributed coding of modalities: The goal of this activity is to explore solutions for joint encoding of multiple modalities. A

natural framework towards this goal is the distributed source-coding paradigm. Distributed source coding (DSC) is a general framework that applies to highly correlated signals that are coded separately and decoded jointly. This framework has been applied to sensor networks, but also to video compression. In the latter application, the motivation is to reduce the complexity of the encoder at the expense of an increase in the decoder's complexity.

In the first period, we have worked on deriving performance bounds and designing a DSC system with a higher number of sources (three correlated sources) for both binary and Gaussian correlation models and designing a practical system based on turbo-codes. This first study has been carried out on theoretical signals verifying given distribution and joint typicality assumptions. One question to be addressed in a context of modalities of different natures concerns the validity of these assumptions for the joint pdf of the modalities considered.

1.10 SIG on Usability

The Usability SIG started almost from scratch and had to identify involved partners, establish a pool of applications for analysis and a common description template for the applications, and then describe each application according to the template and with focus on usability evaluation. A review of state-of-the-art and current practices in usability evaluation has been compiled. The Usability SIG charted usability evaluation issues within SIMILAR theoretical and applicative SIGs. For example, interaction with the fusion-fission SIG is foreseen to integrate usability guidelines in the definition of the architectural framework for multimodal information fusion.

State of the art: The review is necessarily a partial one limited by the partners' areas of expertise. Thus, for example, we have not included any review of systems that can capture smell or systems that can recognise handwriting.

Applications pool: Another large part of the work has focused on establishing a pool of multimodal and natural interactive applications to which the SIG partners had access, establishing a common description template, and describing each application according to the template. The descriptions present the functionality and interface aspects of the applications and have a major focus on usability evaluation.

1.11 SIG on context aware adaptation

The goal of this SIG is to explore all variations in and combinations of the design space:

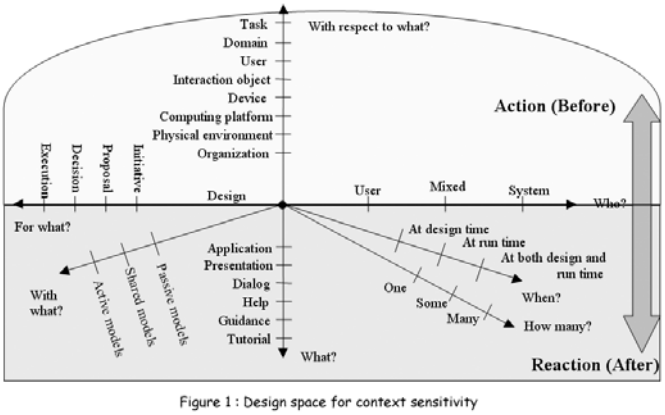


Figure 45: Design space for context sensitivity

This design space is presented like an action-reaction process: Its upper part describes what type of context variation may cause the interface reconfiguration, while the bottom part describes what type of reconfiguration is undertaken as a reaction to the context variation. The SIG will conduct research and development of context-sensitive and context-usable applications by exploring the various combinations of parameters. This task is almost never-ending because of the large number of parameters involved. We thus decided to divide the work into multiple iterations. A first initial task was to design rules. After this initial task, the work on the “intelligent modalities modulator” will be conducted every six months with added modalities.

Research was not that smooth in the beginning as partners realised that deciding upon a common and unified format or for context-specific adaptation rules is impossible because everyone is using a different format, often with different purposes in mind (i.e., for documentation, communication, or automated processing).

We have generated a report that contains an advanced definition of the proposal for a unified format of adaptation rules. The UJF, UCL, and ICI partners have introduced respectively a model based on plasticity of interfaces, guidelines as stored in MetroWeb, and patterns for context-aware adaptation. The question of integrating the format for adaptation rules in the UsiXML (USer Interface eXtensible Markup Language) was also raised. It was discussed that it would be great and coherent to express the adaptation rules in the same specification language that is used to describe the multimodal interfaces. Currently, it is possible to express general rules in UsiXML and its coverage is large enough to accommodate various types of adaptation rules. But it was observed that some desired attributes were missing in UsiXML but found in the other formats.

A prototype version of TransformiXML has been developed. Thanks to this version, it is already possible to test the feasibility of automatic application of adaptation rules on user interfaces specified with UsiXML, without waiting for the first full implementation availability.

1.12 SIG on multimodal analysis and synthesis

The final goal of this SIG is to improve the interpretation of single modalities and to combine the information provided by multiple modalities in order to obtain a more robust and precise interpretation of the interaction scenario.

The various tasks involved in this workpackage have been clustered in three different groups, to wit:

Unimodal signal analysis: Interpretation of speech, facial feature and gesture recognition including within a crowd, handwriting recognition, and brain activity signals interpretation.

Multimodal signal analysis: This area should provide the core research component of the SIG. Basic tools, models, and architectures for the generic analysis of multiple modality signals are being considered. Specially, the collaboration of various groups in tackling the combination of different modalities is to be promoted.

Multimodal synthesis: SIMILAR should concentrate on recent breakthroughs allowing practical use of speech synthesis systems. Furthermore, it should consider modality compensation (a modality is used for more realistic synthesis of another modality) by developing gesture-based speech synthesis (triggering special voice quality or intonation moods as a function of actions to be performed on screen) and speech-based gesture synthesis, mostly for face and sign language synthesis.

In the first year of the SIMILAR network, two big meetings allowed partners to share research results:

May 04 Meeting in Lausanne, Switzerland: Twenty-eight technical presentations covering all technical aspects of the SIG were given by PhD students and researchers involved in SIMILAR. In addition, three tutorial talks were presented:

- "Canonical Correlation Analysis" by Anders Brun, LIU.
- "Modality and multimodality" by Laurence Nigay, INPG.
- "Sequential Monte Carlo methods for audio-visual tracking" by Daniel Gatica-Perez, IDIALAB.

November 04 Meeting in Nice, France: Twenty-nine technical presentations and three tutorial talks were presented. Tutorials were :

- "A Probabilistic Model of Image Mapping with Application to Face Recognition" by F. Perronnin, EURECOM.
- "Coherent Meaningful Blend of Modalities for Natural Behaviour Generation on the example of an Embodied Conversational Character" by Andrea Corradini, NisLab.
- "Distributed Source Coding" by Christine Guillemot, INRIA.

Finally, the achievements were presented in a special session on Multimodal Interfaces at the SPECOM'04 conference in September 2004.

1.13 SIG on medical applications

The three application domains all have their own SIGs. The SIG on medical applications is currently the most important.

The medical applications research was divided into four tasks and several partners have already achieved demonstrable results:

Multimodal Advanced Medical Imaging

- Tensor Imaging (LPI- UVA, LIU, ULPGC, EPFL),
- Work on 3D ultrasound image analysis and classification (LPI-UVA and ULPGC),
- Image segmentation of brain tumours (EPFL).

Advanced Multimodal medical interfaces for learning

- Model of a virtual head for dental training, using data from the Visible Human Project (AUTH).
- Library of organs for a virtual environment (ULPGC and LPI-UVA).

Advanced Multimodal interfaces for operating theatres

- Multimodal display for brachytherapy (EPFL and UCL).
- Project involving ultrasound for arthroscopy therapy (IMPERIAL).

Advanced Multimodal interfaces for medical examination

- Multimodal approach to weaning patients off extubation during surgery (LPI-UVA).
- Algorithms for MRI segmentation and epilepsy assessment (ULPGC).

1.14 SIG on disability applications

Actions carried out in the first year of the SIG for disability applications include many developments:

- A mobile assistant for blind people,
- A haptic environment to access HTML presentations,

- The integration of CyberGrasp and Phantom haptic devices for enhanced haptic access to 3D objects,
- Haptic virtual environments for the training of the visually impaired,
- Videophone for the hearing impaired using MPEG-4,
- Speech driven MPEG-4 facial animation tool for the handicapped,
- Tools for 2D/3D analysis for automated cued speech gesture recognition,
- New approach for hand gesture recognition for sign language teaching,
- Tool for sign language synthesis using MPEG-4 Body Animation Parameters,
- Assistive multimodal system for people who cannot move their hands (based on speech recognition and head tracking),
- New applications of the UPC TTS System for people with disabilities.

Several researchers in the SIG have decided to work within the OpenInterface framework for their future developments. Thus, representatives of the SIG have joined the OpenInterface specifications meetings in order to follow closely the developments in the project.

1.15 SIG on edutainment applications

The aim of this SIG is to provide a variety of input modalities with special emphasis on their applicability in the domain of edutainment and learning applications. In the starting period, due to running co-activities in other research and development projects by some partners, the focus was on producing example applications using the input modalities already present at the

partner's sites and to refine them further on. Objectives of the work package are to "apply multimodal interactions and interfaces to real world, immersive and virtual reality applications to create easy-to-use, intuitive and experience enhancing environments". The undertaken actions are intended to provide (1) first proof-of-concept implementations and (2) reference implementations of what multimodal interfaces and interaction will look like for specific application areas.

The use of digital media, virtual, and augmented reality technology to enhance humans' perceptions of cultural, technical, or artistic assets and to give actors the feeling that they are truly having a ″full experience″ will become increasingly common. The technology offers new channels for communicating, interacting, creating, experiencing, etc. From an educational point of view, if knowledge is transmitted like a real-life game, a journey or an expedition, where actors have to achieve something, it sinks into their minds deeper and faster. Information is memorised more effectively when mediated actively rather than passively. However, most of the current edutainment outcomes are not easy to use and not intuitive enough, because they lack natural interaction paradigms. Recent studies have highlighted a number of usability heuristics to develop such intuitive and seamless interfaces. Among other things, we note that an edutainment system should:

- speak the user's language, with words and concepts familiar to the user, and
- always keep users informed in real time about what is going on.

Virtual sculpturing

Contractor AUTH performed research and integration work aimed at creating a virtual sculpturing tool that operates on 3D volumetric data (voxel-based). This tool has a graphical user interface and in its current form is operated with the mouse. The

user can rotate the volume and use the available sculpturing tools to remove material from the initial volume. Sculpturing (removal of material) is implemented as a series of morphological operations (3D erosions with appropriate structuring elements). The tool keeps track of all removal operations, thereby letting the user "undo" them. Furthermore the user can apply mouse-controlled local smoothing of the volume. Currently the application supports exporting of the surface model in VRML format.

Artistic performance using 3D Body Tracking

Within this task AUTH has investigated the fusion of two tracking concepts, namely, tracking using deformable contours and feature/region based tracking in order to perform accurate contour tracking with initial application in tracking the hand as a deformable object comprising the palm and fingers.

The use of deformable contours in object tracking is a well-known and established technique, especially in hand tracking. Mutual information is well known for its use in image registration. Recently AUTH has proposed the use of mutual information as a similarity metric for feature and region-based tracking. The proposed fusion of these two approaches is achieved by adding a fourth term to the “classic” form of the energy functional that controls the deformation of a deformable contour during tracking. This term is the mutual information between two sets of contour points and can help to avoid contour attraction towards unwanted edges. Mutual information is also used to provide a first estimate of the position of the tracked object in the next frame. The first experiments on hand tracking using this novel approach are very promising. Such a method may potentially be used to enable generation of music through appropriate interpretation of hand or finger movements or for other human-machine interaction tasks.

> "Such a method may potentially be used to enable generation of music through appropriate interpretation of hand or fingers movements."

Contrary to the approach above, ZGDV focused on real-time 3D body motion tracking. This approach was motivated by two experiences conducted in the EU projects PISTE, MEDARPA, and STARMATE.

Decisions were made to use cost-effective infrared tracking for 3D body motion with a restricted set of markers. The development of a real-time 3D body tracking system has started with the prospective goal to enter the gaming market with full-body interaction. A separate recognition module based on a limited set of motion constraints analyses the 3D point sets and assigns the 3D join values to an H-ANIM conform body model in real time.

Pointing interaction

Aligned with research activities in the context of European projects such as artnouveau[2] and art-e-fact[3], ZGDV worked on the implementation of a deviceless interaction using video-based pointing gesture recognition. The interaction was used in a demonstration application called Interactive Museum Exhibit[4]. The application allows exploration of paintings using hand-pointing interaction and was set up with content contributions

---

[2] IST-2001-37865, artnouveau - www.zgdv.de/zgdv/departments/z2/Z2Projects/artnouveau

[3] IST-2001-37924, art-e-fact - www.zgdv.de/zgdv/departments/z2/Z2Projects/art-e-fact

[4] Interactive Museum Exhibit - www.zgdv.de/zgdv/departments/z2/Z2Projects/IME/index_html_en

from the Vienna Academy of Fine Arts, a partner in the art-e-fact project.

The initial version of the pointing gesture recognition used a camera set-up (left-view/top-view mounting) that allowed computation of the pointing direction without initial calibration of the cameras. However, due to certain disadvantages, this set-up was replaced with a fully top-view camera set-up, reducing potential occlusion problems and constraints regarding light conditions. The pointing gesture is assumed to be “static”, although the recognition can handle individual characteristics of pointing gestures. Currently, the fingertip is identified in both camera images using appropriate image processing algorithms and the pointing direction is estimated based on image analysis results.

The 2D/3D pointing information is used to enable x-ray views on an image, for example. At the pointing hotspot, parts of an x-ray image become visible, depicting underlying analysis information, which shows that the original canvas of the painting was already used before. In addition, pointing gesture recognition can be configured to trigger specific events if the user points to a certain area of interest. These events activate multimodal output such as spoken text (using an integrated text synthesis module) either with or without virtual character presence, additional visual augmentations, video, audio, and animations. The application has been further extended so that it isnot only applicable to painting exploration, but can handle complex 3D models using VR technology as well.

INPG-LIS also worked on the topic of video-based interaction for edutainment or education applications using immersive environments. They have developed an algorithm to segment the arm of a moving person in front of a camera. Unlike the approach described above, in which the user takes a designated

position in front of displaying canvas, this approach handles moving persons.

Multimodal Speaker Identification

Biometric person recognition technologies include recognition of faces, fingerprints, voice, signature strokes, iris and retina scans, and gait. Person recognition in general encompasses two different, but closely related, tasks: identification and verification. The former refers to identification of a person from her/his biometric data from a set of candidates, while the latter refers to verification of a person's biometric data; hence the output is binary. It is generally agreed that no single biometric technology will meet the needs of all potential recognition applications. Relatively little work on the fusion of the results of various biometric technologies has been reported in the literature. A particular problem of interest in multimodal identification is the speaker identification problem using both audio (voice) and video signals.

A multimodal open-set speaker identification system using audio signals, face, and lip motion is being considered. For fusion of multiple modalities, we propose a new rule that favours reliable modality combination through a cascade of classifiers. The order of the classifiers in the cascade is adaptively determined based on the reliability of each modality combination. The proposed adaptive rule is more robust in the presence of unreliable modalities, and outperforms the hard-level max rule and soft-level weighted summation rule. The lead partner in this project is KOC (Turkey).

Surveillance and assistance in learning environments

This task is focusing on head interaction. In surveillance and assistance, explicit head movements, gaze tracking, speech, and

implicit movements such as facial expressions provide essential information about the user to enable assistance systems to adapt appropriately and automatically.

INPG-LIS focused its research on multimodal interaction and in particular on the head. They have developed a frequencies method for rigid 2D head motion analysis based on video allowing the automatic detection and interpretation of head rotation. Head oscillations typical in human expressions of denial or affirmation using the head (head shaking and nodding) are easily detected.

Furthermore, we have implemented a facial expression analysis system based on the fusion of low level information about features' contours (mouth, eyes, and brows). Data fusion is performed by the use of belief theory. INPG-LIS set up an interactive video test bed. It consists of two monocular digital colour cameras, two fast workstations, and diffuse lights. Several visual servo-ing algorithms have been used to command the camera tilt and zoom. This test bed has been used to test in real conditions and accelerate several tracking and recognition algorithms.

1.16 Communication between partners

Communications between partners essentially took place by e-mail (more than 90%), but also by phone (9%) and fax (less than 1%). Very few documents were physically transmitted (books and official signatures mainly). Since 2004, the partners have begun using the free "Voice over IP" communication tool called "Skype" (see www.skype.com). This tool is developed in Europe and is very easy to use on a PC with cheap speakers and a microphone (investment per workstation typically under 30

euros). To give an idea of the volume of exchanges during 2004, the SIMILAR coordination office received 2,360 e-mails and 450 phone calls, sent 1,890 e-mails (with more than 500 sent to multiple partners), and made 300 phone calls. As Skype communication is free, the traffic has not been logged, but is estimated to be over 15 hours per month.

Figure 46: General Assembly Meeting

Steering Committee Meeting

Communication between partners inside and between the various SIGs has not been logged but has been evaluated at around 25,000 e-mail messages exchanged, 4,000 of them with attached documents.

1.17 Submission of European research projects

The collaboration between SIMILAR partners was recently concretised with the setting up and submission of two research projects to FP6 IST Call 4, for the objective "Integrated biomedical information for better health" on 22 March 2005.

The first one is an Integrated Project (IP) called THERAGIS (standing for "from THEranostic RAdiotherapy to Global Information System"). Its purpose is to develop a Clinical Information System integrating all levels of patient data, from system to gene, and a Treatment Workstation dedicated to cancer treatment (radiotherapy and brachytherapy), especially for head and neck, lung, and prostate cancers, comprising among others a collaborative tool allowing the fostering of various expertise: medical doctors, radiotherapists, oncologists, molecular biologists, etc. The design of these tools will obviously integrate multimodal aspects.

The THERAGIS consortium consists of thirteen participants covering academic, clinical and industrial fields: universities, hospitals, companies and the European Organisation for the Research and Treatment of Cancer.

All universities collaborating in THERAGIS except the teaching hospitals, i.e., Imperial College, UCL, EPFL, CLIPS, UPC, LIU are SIMILAR members. What is more, two of the four companies involved in the project belong to SIMILAR's Industrial Advisory Board.

The second research project submitted pour IST Call 4 is a Specific Targeted Research Project (STREP) called SAPHIR (Semantic-based Annotations based on Propagation of Human InteRactions). It is dedicated to image-based data mining for 3D objects, among others for medical applications. This will obviously require the definition of a new ontology, as well as the development of visualisation tools, which will certainly include multimodal aspects.

Four of the six SAPHIR partners (i.e., UCL, UPC, ZGDV, and MULTITEL) are members of SIMILAR.

These projects constitute a good example of the level of communication and co-operation inside the SIMILAR network. Other projects have also been jointly submitted by several SIMILAR members for national and/or regional funding.

Other submissions of European research projects (either IPs either STREPs) by SIMILAR members, potentially with partners from outside the Network of Excellence, are being envisioned for future calls.

Part 3: Dreams

Chapter 9: Introduction

Sandra Daussogne (UCL, Belgium)

9.1. The “10-10-10”

Chapters 10 to 12 are the result of the fruitful collaboration of the lab heads attending the senior Lausanne fusion-fission meeting, namely, Michel Barlaud, Niels Ole Bernsen, Joëlle Coutaz, Jean-Luc Dugelay, Benoît Macq, Marcos Martín, Laurence Nigay, Fabio Paternò, Michael Strinzis, Murat Tekalp, Jean-Philippe Thiran, Dimitrios Tzovaras, Jean Vanderdonckt, and Slava Voloshynovskiy.

The SIMILAR senior fusion-fission meeting focused on the concepts of fusion and fission, which are obviously central to multimodal interfaces. But, from a wider point of view, it was also an occasion for those attending to compare their opinions on the past, present, and future of multimodal interfaces.

What has emerged from these discussions are what we have called the “three times ten” or “10-10-10”, i.e., 10 research achievements, 10 practical achievements, and 10 grand challenges.

These topics cover HCI as well as DSP, and for the grand challenges we have tried and hope to combine the two as much as possible.

Because of the wide range of application domains covered, from pure software engineering to medical, and because of their synthetic character, these lists are also deemed to constitute a valuable ground for discussion between HCI and DSP, within SIMILAR but also with experts outside the NoE.

Commenting the topics, ranking them, or suggesting other ones, and explaining the grounds for our choices is a very simple way to launch discussion on respective views on the past, present, and future of multimodal interfaces.

9.2. Interviews

SIMILAR gathers European expertise in multimodal interfaces, from HCI as well as DSP, and its members are involved in several research projects at the national, regional, and European level, where they notably collaborate with other people from outside SIMILAR. The network is thus aware that research in the domain is also ongoing outside the NoE, and its ambition is to establish and maintain strong links with outside expertise in order to keep abreast of the state of the art in the field and to disseminate its own progress and results as much as possible.

Conferences, publications, summer schools and this discussion about the « 10-10-10 », are ways to achieve this goal. But another way is to interview different people from the SIMILAR community but outside SIMILAR as well and collect their answers to the same set of questions. This option was chosen to deal with visions of the future in particular, with interviewees coming from both academia and industry.

This collection of interviews, which is destined to get richer with time, will allow the fruitful comparison of many different points of view, and many open space for new collaborations.

These interviews (taking the form of face-to-face, phone, and e-mail interviews) were conducted by Sandra Daussogne and Benoit Macq between May 2004 and June 2005.

Chapter 10: Ten major achievements

Lausanne Senior Meeting

10.1. Reliability-ordered cascade

The multimodal signal community has made major progress on reliability ordered cascade for fusion of multimodal signals. Murat Tekalp's team contributed in this domain with biometry applications in mind. Different strategies are possible: In the so-called "early integration" modalities are fused at data or feature level, whereas in "late integration" decisions or scores resulting from each unimodal classification are combined to give the final conclusion. This latter strategy is also referred to as decision or opinion fusion and is effective especially when the contributing modalities are uncorrelated and thus the resulting partial decisions are statistically independent. Early integration techniques, on the other hand, , can be favoured if they are used properly and a pair of modalities is highly correlated, as in the fusion of audio and lip movement. Multimodal decision fusion can also be viewed from a broader perspective as a way of combining classifiers, which is a well-studied problem in pattern recognition. The main motivation of Murat Tekalp and his team's work is to compensate for possible misclassification errors of a certain classifier with other available classifiers and to end up

with a more reliable overall decision. Misclassification errors are in general inevitable due to numerous factors, such as environmental noise, measurement and modelling errors, or the time-varying characteristics of signals.

Recent work uses decision-level fusion for the verification problem, where scores resulting from each classifier are concatenated to form a feature vector, which is then fed into another classifier, e.g. a median radial basis function (MRBF) network or support vector machines and Bayesian classifier.

The work of Murat Tekalp and co-workers presents a new multimodality fusion strategy where some of the modalities might be corrupted by measurement noise and/or modelling errors. The basic idea is that a single highly reliable modality alone may sometimes yield a correct decision, whereas its linear fusion with some other less reliable modality may give incorrect results. On other occasions, results obtained by the fusion of two modalities may outperform those obtained from each modality alone. Hence, their proposed scheme considers all possible linearly fused modality combinations (including single modalities) with their corresponding reliability measures and strives to maximise the benefit of multimodal fusion so that the upper bound for the system error rate becomes the expected occurrence rate of the cases where all classifier combinations fail. Thus, a critical feature of their system is to be able to assess each modality classifier adaptively with a reliability measure.

10.2. Mutual information in multimodal SP

10.2.1. Mutual information

In probability theory and information theory, the mutual information of two random variables is a quantity that measures the independence of the two variables.

If X and Y are discrete random variables and $f(x,y)$ is the value of their joint probability distribution at (x,y), and $f(x)$ and $f(y)$ are the marginal distributions of X and Y, respectively, then:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} f(x,y) \log \frac{f(x,y)}{f(x)f(y)}$$

is the mutual information (MI) of X and Y.

Figure 47: Mutual information formula

Informally, mutual information measures the information of X that is shared by Y. If X and Y are independent, then X contains no information about Y and vice versa, so their mutual information is zero.

10.2.2. Mutual information in multimodal signal processing

Most of the time the various modalities that are processed consist of signals that are more or less related to each other. It was thus obvious to use the information-theoretic notion of mutual information.

This is obviously the case when several medical images of the same area acquired through different modalities are to be registered. For two images, mutual information is a measure of how well one image explains the other, or vice versa, and is assumed to be maximal if the images are geometrically aligned. Maximisation of mutual information is a very general and powerful criterion because no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved.

Another example of the use of mutual information is the combination of audio and image or video as the voice modality will have an impact on at least one other modality, namely, the image of the face, as the person will open his/her mouth to emit sounds.

Figure 48: MI between image and audio:
The maximum clearly lies at the speaker' mouth

10.3. Multimodal emotion detection

10.3.1. Introduction

Human communication is of course composed of verbal or explicit discourse, but it also includes many implicit components, such as emotions. Several studies depicts the importance of non-verbal behaviour in human-human communication. These studies show that more than half of semantic exchanges happen through non-verbal behaviour (gestures, facial expressions, postures, etc.)

The ability to understand such emotions is then highly desirable for the computer in HCI, especially in applications where the computer has to play the role of a teacher, helper, or companion.

Emotion recognition systems will allow natural interactions between computer and users, as well as learning users' preferences.

Information about emotions can be acquired through facial expressions, but also through language, gestures, and other physiological components (clamminess of hands, etc.)

It has been proven that humans behave purely multi-modally when recognising emotions. This area was thus an obvious choice for multimodality research.

10.3.2. Achievements

Many studies were conducted about the various physiological signs of emotions, with the face in first position. As a result of these studies, six universal face expressions were documents, namely, happiness, sadness, anger, fear, surprise, and disgust.

Other studies were conducted concerning voice, gestures, and so on.

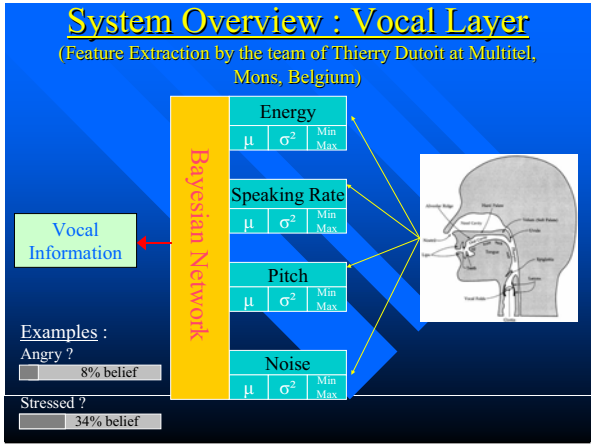


Figure 49: Emotion recognition through the voice

One robust multimodal interface has been set up for the recognition of emotions. It effects a fusion between face and prosody, i.e., the melody or musical component of the spoken voice.

10.3.3. Future research work

Research to improve robustness and to extend the scope of detected emotions is ongoing.

This work either integrates other modalities, such as gesture, or uses other fusion methods that are deemed to be more robust.

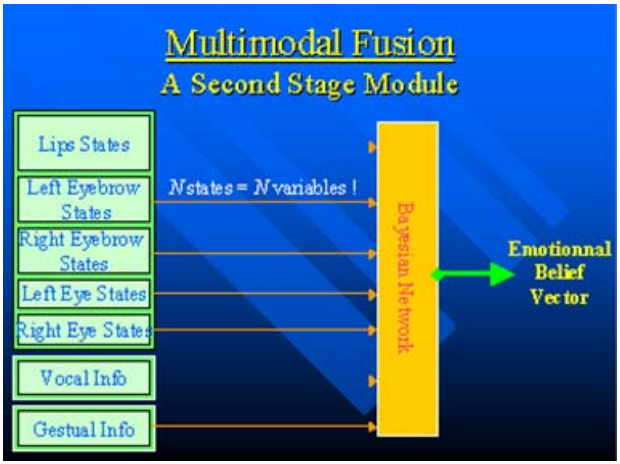


Figure 50: A scheme of ongoing research in emotion recognition

10.4. Multimodal biometrics

10.4.1. Introduction

Biometrics deals with authentication and identification based on a person's characteristics. Each person has a unique anatomy and behaviour.

Applications are numerous. For example, a multi-player computer game can interact naturally with the players by recognising them. Other examples include access control to

restricted secure area, prisoner monitoring, passenger identification in airports, and secure virtual private network access. Access can be controlled using a person's face image, iris scan, and/or a few seconds of speech.

Our brains recognise and check our peers' identities hundreds of times a day with very little error. This task is performed using visual information, for example by looking at the person's face or gait, or auditory information, for example by listening to a person's voice, or both.

This shows that our brains are able to extract the relevant features from the information flow that we perceive. These features are:

(i) discriminative;

(ii) compact, so as to be memorised easily; and

(iii) retrieved very quickly so as to be matched against new incoming data to recognise.

If we could determine these features and the accompanying recognition process, the identity check or recognition task could be automated, opening the doors to dozens of applications in data protection, security, and entertainment. Given two face images, we must determine automatically whether they are two images of the same person or of different persons. The field of automatic personal identity verification and recognition using human specific characteristics is called biometrics. The problem is difficult due to many factors, such as variability of facial appearance, sensitivity to noise, template aging, etc..

10.4.2. Different Fusion Strategies

We can overcome some of these difficulties by combining different information sources for the classification/recognition task. Strategies for combining the different information sources, i.e. fusion strategies, have been proposed in order to improve the verification's accuracy. As all algorithms operate on the same modality (face modality in this case), this fusion type is referred

to as intramodal fusion. This is an example of fusion of different processing algorithms working on the same signal.

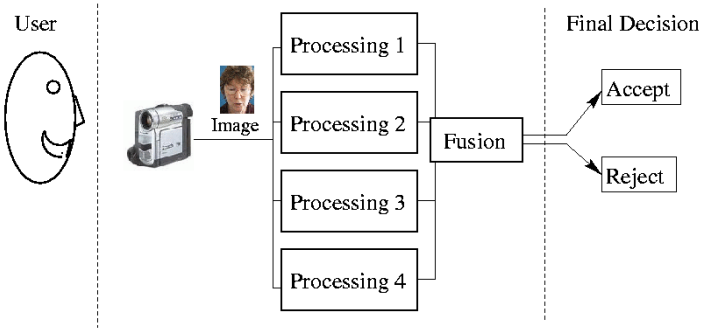


Figure 51: Intramodal fusion: One signal, several processing steps

Another fusion aspect that can be considered is the fusion of confidences obtained sequentially on several video frames of the same person's face. Instead of making the decision based on a single image, several images are taken of the person whose identity is to be verified. This type of fusion is referred to as **Multi-frame fusion**.

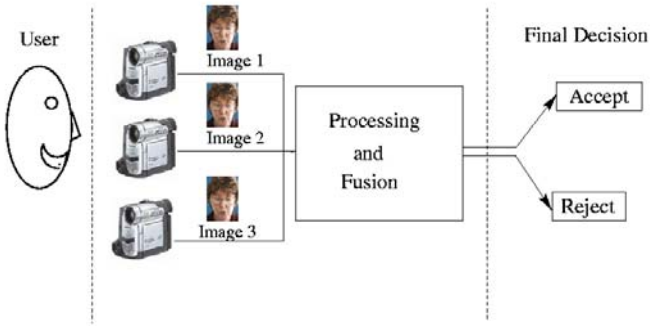


Figure 52: Multi-sensor fusion. Several cameras capture the same modality but at different angles from the subject.

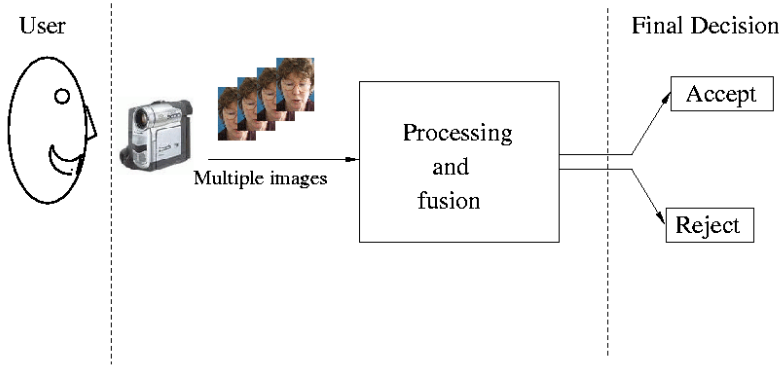


Figure 53: Multiframe fusion: Several images, one processing step

The **Multimodal fusion** of face and speaker verification algorithms[1] illustrates the case of different sensors capturing different physical phenomena related to the same state of nature.

In this case, the speech and face of the same subject are recorded and processed by different algorithms, which yield separate opinions. These two outputs (opinions) are then conciliated at the fusion stage.

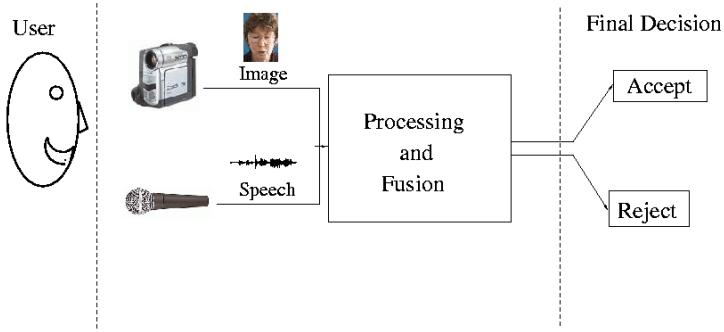


Figure 54: Multimodal fusion Different sensors, such as a camera and a microphone, record different physical phenomena related to the same identity.

Much theoretical and experimental evidence shows that the best accuracy improvement is obtained with multimodal fusion. The main reason for this is that the noise perturbing the biometric signals is much less correlated (or even independent) for two different modalities in comparison with the other fusion schemes presented here.

10.5. Multimodal medical image registration

10.5.1. Image registration

Image registration can be defined as the spatial alignment transformation (mapping) used to match two or more pictures. The registration is classified according to the origin of the differences existing between the pictures to match.

It is thus:

- multimodal if the ways of acquiring the pictures are different (e.g., various sensors),
- temporal if the time varies (e.g., before and after surgery),
- viewpoint if registration is used to infer 3D information from images in which either the camera or the objects in the scene have moved,
- template for model-based object recognition.

Medical image registration is used for three main applications.

The first one, on which we shall focus hereafter, deals with different images of the same individual acquired by different means.

The other two, which entail a single modality, treat either pictures from multiples patients or images from the same patient at different times.

10.5.2. Multimodal medical image registration

The need for multimodal medical image registration comes from the rather extended number of medical imaging technologies, each of which has its own advantages:

- some primarily depict anatomy/morphology (CT, MR, etc.), whereas some primarily depict information about the metabolism of the underlying anatomy (PET, IMRI, EGG, etc.). This is the basic distinction between anatomical and functional imaging,

- within the anatomical modalities, some – for example, X-rays - are dedicated to bones, whereas others – such as US - give images of soft tissues.

Image registration lets one integrate different images into one representation such that the complementary information can be accessed more easily and accurately. Multimodal images of the same person or of different persons generally differ by local geometric differences. Consequently, non-rigid or elastic transformations are required to map such images onto one coordinate system.

The resulting aligned images give medical professionals a much better view of the imaged region. Medical image registration has been used in the diagnosis of breast cancer, cardiac studies, and a variety of neurological disorders, including brain tumours. It also allows better surgery planning and simulation as well as intra-operative navigation.

Many registration algorithms exist.

To perform the matching, several various criteria and systems (neural networks, statistical inferences, maximisation of mutual information, etc.) are used in several domains, from spatial to wavelets.

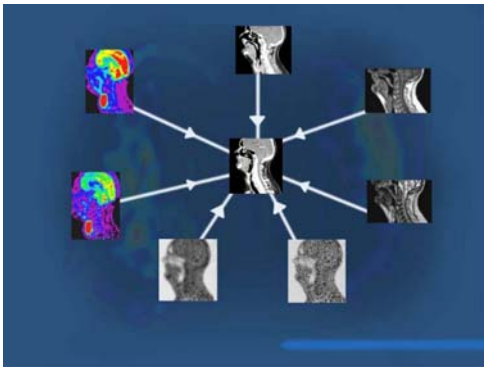


Figure 55: Brain images registration

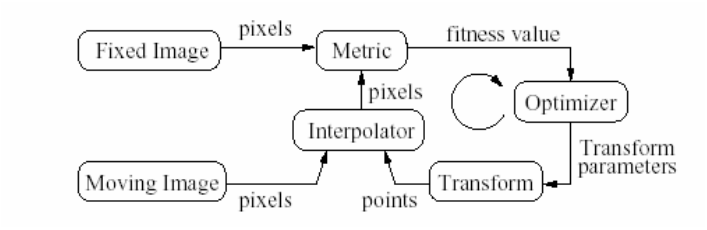


Figure 56: One possible image registration scheme

10.6. ICARE

10.6.1. Introduction: The CARE properties

The CARE properties - the Complementarity, Assignment, Redundancy, and Equivalence that may occur between the interaction techniques available in a multimodal user interface - have been designed as a simple way of characterising and assessing aspects of multimodal interaction.

Complementarity: Modalities of a set M must be used in a complementary way to reach state s' from state s within a temporal window; if all of them must be used to reach s' from s, i.e., none of them taken individually can cover the target state.

To express this adequately, we need to extend the notion of reachability to encompass sets of modalities: REACH(s,M,s') means that state s' can be reached from state s using the modalities in set M.

$$
\begin{aligned}
&Complementarity\ (s, M, s', tw) \\
&\square\ \ (Card(M) > 1) \wedge (Duration(tw) \neq \infty) \\
&\wedge (\forall M' \in \mathbf{P}M\ (M'\ \ M \Rightarrow \neg REACH\ (s, M', s'))) \\
&\wedge REACH\ (s, M, s') \wedge (\forall m \in M \cdot \exists t \in tw \cdot Active(m, t))
\end{aligned}
$$

Figure 57: Complementarity formula

Assignment: Modality m is assigned in state s to reach s' if no other modality is used to reach s' from s. In contrast to equivalence, assignment expresses the absence of choice: either there is no choice at all to get from one state to another, or there is a choice but the agent always opts for the same modality to get between these two states. Thus we can define two types of assignment:

$$
\begin{aligned}
&StrictAssignment\ (s, m, s') \square\ \ Reach\ (s, m, s') \\
&\wedge\ (\forall m' \in M\ (m' \neq m \Rightarrow \neg Reach\ (s, m', s'))) \\
&AgentAssignment\ (s, m, M, s') \square\ \ (Card(M) > 1) \\
&\wedge\ (\forall m' \in M\ Reach\ (s, m', s') \wedge (Pick\ (s, m', s') \\
&\Rightarrow m' = m)\ )
\end{aligned}
$$

Figure 58: A: assignment formula

Redundancy: Modalities of a set M are used redundantly to reach state s' from state s, if they have the same expressive power (they are equivalent) and if all of them are used within the same temporal window, tw. In other words, the agent shows repetitive behaviour without increasing its expressive power:

$$
\begin{aligned}
&Redundancy\ (s, M, s', tw) \square\ \ Equivalence\ (s, M, s') \\
&\wedge\ (\forall m \in M \cdot \exists t \in tw \cdot Active(m, t))
\end{aligned}
$$

Figure 59: Redundancy formula

Equivalence: Modalities of set M are equivalent for reaching s' from s if it is necessary and sufficient to use any one of the modalities. M is assumed to contain at least two modalities. More formally:

$$Equivalence\ (s, M, s') \Leftrightarrow (Card(M) > 1) \wedge (\forall m \in M\ Reach\ (s, m, s'))$$

Figure 60: Equivalence formula

Equivalence expresses the availability of choice between multiple modalities but does not impose any form of temporal constraint on them.

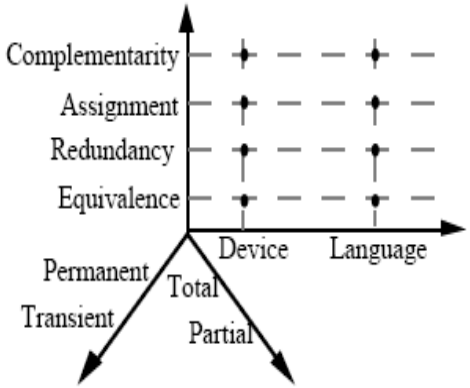


Figure 61: A framework to characterise multi-feature user interfaces with relations between interaction languages, physical devices, and tasks

10.6.2. ICARE: What is it?

ICARE is a design environment for specifying and designing multimodal user interfaces based on the CARE properties (Complementarity, Assignment, Redundancy, and Equivalence).

Systems support multiple interaction techniques such as the synergistic use of several modalities (speech, gesture, gaze, localisation of a user, etc.). The flexibility that they offer results in increased complexity that current software development tools do not address appropriately. ICARE (Interaction CARE - Complementarity Assignment, Redundancy and Equivalence-) is a component-based approach that allows the easy and rapid development of multimodal interfaces. ICARE is concentrated on input (i.e., from the user to the system), although our model holds for output as well. Nevertheless, we have not tested our approach for output so far.

The ICARE conceptual model includes two kinds of ICARE component:

1. **Elementary components:** Such components are building blocks useful for defining a modality. Two types of ICARE elementary components are defined: Device components and Interaction Language components. We re-use the definition of a modality as the coupling of a physical device d with an interaction language L: <d, L>. A *physical device* is an artefact of the system that acquires (input device) information. Examples of devices include the mouse, microphone, GPS, and magnetometer. An *interaction language* defines a set of well-formed expressions (i.e., a conventional assembly of symbols) that convey meaning. The generation of a symbol, or a set of symbols, results from actions on physical devices. Examples of interaction languages include pseudo-natural language, direct manipulation, and localisation. An *interaction modality* such as speech input is then described as the couple <microphone, pseudo natural language NL>, where NL is defined by a specific grammar. Similarly graphic input is described in terms of <mouse, direct manipulation>.

2. **Composition components:** Such components describe combined usages of modalities and therefore enable us to define new composed modalities. The ICARE composition components are defined on the basis of the four CARE properties - Complementarity, Assignment, Redundancy, and Equivalence -

that may occur between the modalities available in a multimodal user interface. Composition components can connect from two to n components and are not dependent on a particular modality.

10.6.3. Elementary Components

Elementary components are dedicated to interaction modalities. Based on the definition of an interaction modality, we identify two types of elementary ICARE components, namely Device and Interaction Language components.

An **ICARE Device** component represents a supplemental layer of the physical device driver. For example, the mouse Device component abstracts the data provided by the mouse driver such as button pressed/released and movement. Likewise a microphone Device component abstracts the captured signal into a recognised utterance while another microphone Device component abstracts the captured signal into a level of noise. All ICARE Device components also enrich the raw data from the device driver by adding information that includes the device's operating status, time-stamp, confidence factor of the produced data, and a description of the device in terms of human manipulation (passive/active modalities, human actions involved, and physical location of these actions). An ICARE Device component is then linked to a listener component, an ICARE Interaction Language component, in order to form an *interaction modality*.

An **ICARE Interaction** Language component corresponds to the logical level of an interaction modality. For example, an Interaction Language component abstracts the data from a mouse Device component into commands such as the selection of a menu option. Similarly, another Interaction Language component (NL component) abstracts a set of characters from a microphone Device component (recognised utterance) or from a keyboard Device component into a command. A third example, shown in Figure 62, corresponds to a passive modality: the 3D Location component that abstracts data from a localisation sensor (e.g., GPS) Device component into a user's location

expressed in a given coordinate system. These three examples of Interaction Language components underline the fact that such components may need to rely on an external description of well-formed expressions to be obtained. Indeed, to abstract data from the mouse into commands, a description of the graphical interface is required. Likewise, NL recognition requires a description of the pseudo Natural Language to be recognised (NL grammar). Finally, the 3D Location component may require a description of the environment of the user in order to produce an event such as <the user is entering a particular room>. Just as Device components are dependent on the underlying physical devices, Interaction Language components are dependent on a class of Device components that can produce the required inputs. For example, an NL component requires a set of characters as inputs that can for example be produced by a microphone Device component or a keyboard Device component. Finally, like ICARE Device components, ICARE Interaction Language components also enrich the data by adding generic information that include the time-stamp as well as a confidence factor of the produced data.

Device and Interaction Language components are the building blocks for defining modalities. The designer can then combine these components in order to specify a new composed modality, in other words, a combined usage of several modalities.

10.6.4. Composition Components

The CARE properties characterise the different usages of multiple modalities. Based on the CARE properties, we define three composition components: t: Complementarity component, Redundancy component, Equivalence component, and Redundancy/Equivalence component. Assignment and Equivalence are not modelled as components in our ICARE model. Indeed, as shown in Figure 62, an assignment is represented by a single link between two components. An ICARE component "Magnetometer" linked to a single component "3d orientation" implies that "Magnetometer" is

assigned to "3d orientation". As for Assignment, Equivalence is not modelled as a component. As shown in Figure 63, when several components (2 to n components) are linked to the same component, they are equivalent. In our previous multimodal systems developed using ICARE components, we explicitly used an Equivalence component that had no functional role (no treatment except defining a new time-stamp for the data) but constituted an aid while manually assembling components. Using our platform under construction that will allow the user to graphically assemble ICARE components, such Equivalence component has no more utility.

In Figure 62 we present an example of ICARE specifications that includes Complementarity components. Let us consider the Complementarity-1 component. In order to compute the location and orientation of the user that is required by the application, two passive modalities are used in a complementary way. The Complementarity-1 component of figure 62 defines a customisable temporal window for merging data received by the two Interaction Language components (respectively orientation in radians and location as latitude/longitude in WGS84 normalization and altitude in meters).

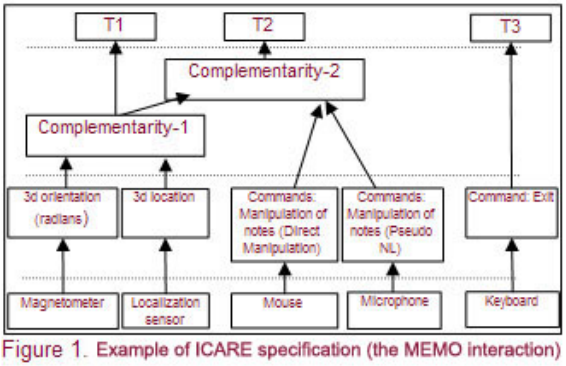


Figure 62. Example of ICARE specification

As ICARE elementary components, ICARE composition components enrich the data by adding generic information that

includes the time-stamp and a confidence factor of the produced combined data. In addition composition components include parameters that the designer can fix for customizing the composition mechanism.

10.6.5. ICARE Platform

The ICARE platform enables the designer to manipulate and assemble ICARE software components graphically in order to specify the multimodal interaction dedicated to a given task of the interactive system under development. From this specification, the code is generated automatically. To understand the scope of our ICARE platform fully we show in Figure 63 where the automatically generated code is located within the complete code of the interactive system structured along the ARCH software architectural model.

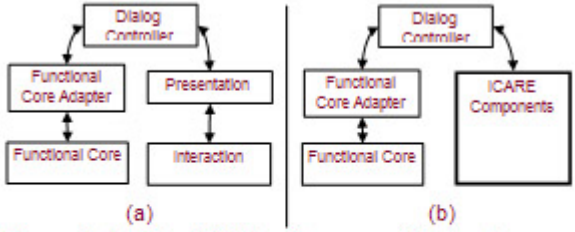


Figure 2. (a) The ARCH software architectural model.
(b) ICARE components within an ARCH software architecture.

Figure 63

The originality of the ICARE platform lies in the fact that it is for designers, not developers. Indeed the user of the ICARE platform selects the modalities and specifies the combination of modalities in terms of the CARE ergonomic properties, all by assembling software components graphically without knowing the details of the components' code. From this high level

specification, the code of the input multimodal UI is then generated.

Figure 64 presents a sketch of the ICARE platform's user interface. It contains a palette of components, an editing zone for assembling the selected components, and a customisation panel for setting the parameters of the components. Although the complete ICARE platform is not yet available, we have already designed and developed several components, including modality components, as well as combination components in order to validate our approach.

By assembling these components manually, we have developed several multimodal systems. The following section describes these ICARE components that will in the near future be manipulated graphically in the ICARE platform.

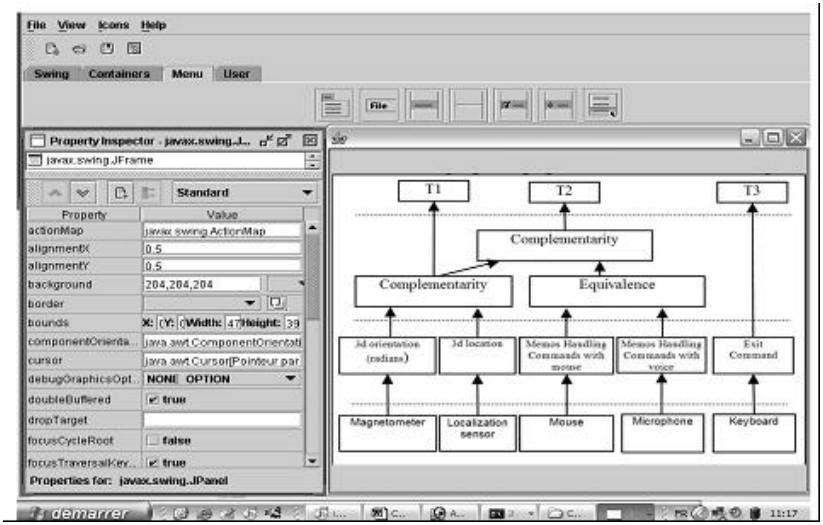


Figure 64. Sketch of the graphical ICARE platform.

10.7. Cameleon

The CAMELEON (Context Aware Modelling for Enabling and Leveraging Effective interactiON) Reference Framework is intended to support the development of context-sensitive user interfaces in a model-based approach. It describes models at four abstraction levels (Figure 65) from task specification to the running interface:

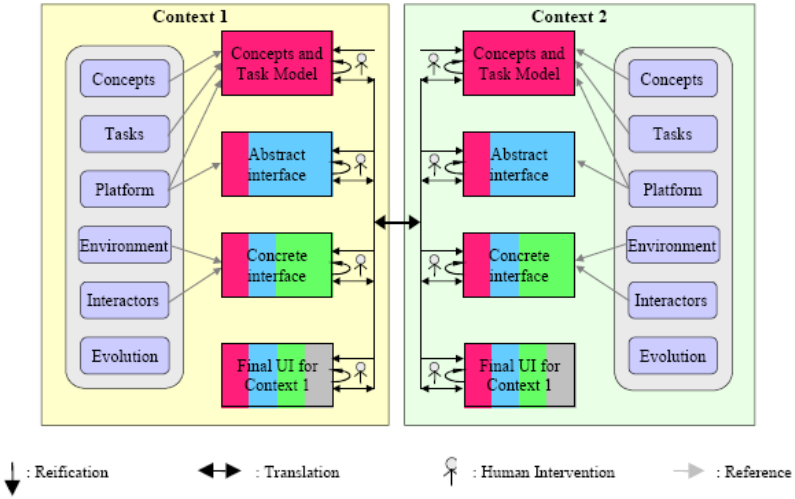


Figure 65. Reference framework of plastic interactive systems.

- The **Tasks and Concepts** level describes the interactive systems' specifications in terms of the user tasks to be carried out and the domain objects manipulated by these tasks.

- The **Abstract User Interface** (AUI) is an expression of the UI in terms of presentation units, independently of which interactors are available. A presentation unit is a presentation environment (e.g., a window or panel) that supports the execution of a set of logically connected tasks.

- The **Concrete User Interface** (CUI) is an expression of the UI in terms of abstract interactors and their position. The concrete UI is still only a mock-up in the development environment. It can be modified by the designer.

- The **Final User Interface** (FUI) is generated from a concrete UI expressed in the source code of any programming language or mark-up language (e.g. Java or HTML). It can then be interpreted or compiled.

The main contribution of the reference framework is its capability to be instantiated in many ways. This clearly characterises the functional coverage of current tools and makes it possible to make comparisons and identify requirements for future tools. Examples of tools developed in the CAMELEON reference framework are TERESA and Vaquita/ReversiXML. TERESA generates UIs for multiple devices from a single task model. The model is filtered to produce a platform specific task model. This task model is further transformed into an Abstract User Interface (AUI). Vaquita/ReversiXML extracts an abstract description from a UI's code and generates the code of new UIs adapted to the target platform(s).

Drawing upon the reference framework, CAMELEON-RT (Run Time) is the software infrastructure for accommodating the dynamic adaptation of context-sensitive user interfaces in heterogenous information spaces.

The architectural model is organised in three levels of abstraction:

The **Interactive Systems layer** consists of the interactive applications that users currently run in the information space.

The **DMP-middleware layer** provides services for modelling the physical space (context infrastructure), supporting the creation of dynamic heterogeneous clusters of resources, and adapting the UI when distribution and migration occur.

The **platform layer** groups the hardware and legacy operating system(s) of an interactive space. The hardware includes surfaces and instruments, computing and communication facilities, and sensors and actuators.

Figure 66 shows CamNote, which is a slide viewer implementing the CAMELEON-RT architecture and running on a dynamic heterogeneous platform. This platform may range from a single PC to a cluster composed of a PC and a PDA. The viewer switches between the platform configurations automatically according to the available devices. The user interfaces are also reshuffled automatically.

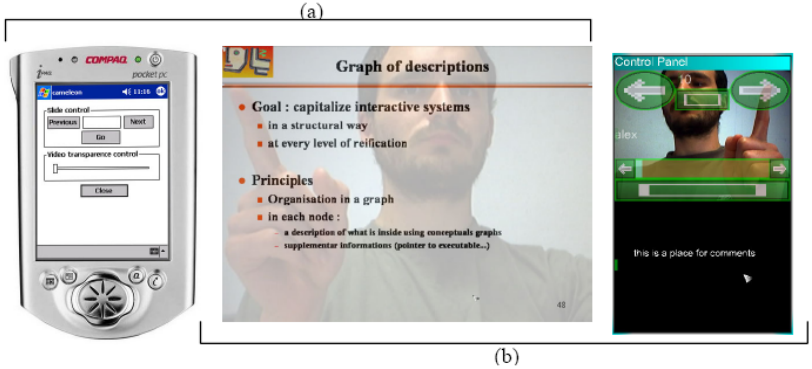


Figure 66. (a) The user interface of CamNote when distributed on a PC and a PocketPC screens; (b) the control panel when displayed on the PC screen.

10.8. Modality Theory

To achieve at least part of the understanding needed, it appears that the following objectives defining the research agenda of Modality Theory (Bernsen 1993) should be pursued:

(1) To establish an exhaustive taxonomy and systematic analysis of the unimodal modalities that go into the creation of multimodal *output* representations of information.

(2) To establish an exhaustive taxonomy and systematic analysis of the unimodal modalities that go into the creation of multimodal *input* representations of information. Together with Step (1) above, this will provide sound foundations for describing and analysing any particular system for interactive representation and exchange of information.

(3) To establish principles for how to legitimately combine different unimodal output modalities, input modalities, and input/output modalities for usable representation and exchange of information.

(4) To develop a methodology for applying the results of Steps (1) – (3) above to the early design analysis of how to map from the requirements specification of some application to a usable selection of input/output modalities.

(5) To use results in building (possibly automated) practical interaction design support tools.

The research agenda of Modality Theory thus addresses the following general problem: Given any particular set of information that needs to be exchanged between user and system during task performance in context, identify the input/output modalities that constitute an optimal solution to the representation and exchange of that information. This is a hard problem for two reasons. Firstly, already at the level of theory there are a considerable number of unimodal modalities to consider whose combinatorics, therefore, is quite staggering.

Secondly, when it comes to applying the theory in development practice, the context of use of a particular application must be taken thoroughly into account in terms of task, intended user group(s), work environment, relevant performance and learning parameters, human cognitive properties, etc.

A particular modality is not simply good or bad at representing a certain type of information – its aptness for a particular application very much depends on the context. This adds to the combinatorics generated by the theory to form an open-ended space of possibilities for consideration by the developer, a space which, furthermore, remains poorly mastered despite decades of HCI (Human to Computer Interaction)/HHSI (Human-Human-System Interaction) research, primarily because such is the nature of engineering as opposed to abstract theory.

The space of unimodal output representations can be carved up at different levels of abstraction. We have seen that already above, in fact, because the three media of graphics, acoustics, and haptics may be viewed as a very general way of structuring the space of unimodal output representations. What will be proposed in the following is a downwards extensible, hierarchical generative taxonomy of unimodal output modalities that at present has four levels: a super level, a generic level, an atomic level, and a sub-atomic level. In terms of the generative steps to be made, the generic level comes first. Thus, the taxonomy is based on a limited set of well-understood generic unimodal modalities. The generic modalities are generated in turn from sets of basic properties.

We generate the generic-level unimodal output modalities from a small set of *basic properties* that serve as a robust way to distinguish modalities from one another within the taxonomy. The properties are: *linguistic*/*non-linguistic, analogue*/*non-analogue, arbitrary*/*non-arbitrary,* and *static–dynamic.* In addition, a distinction is made between the physical *media of expression* of graphics, acoustics, and haptics, each of which is characterised by a very different set of perceptual qualities (visual, auditory and

tactile, respectively). These media determine the scope of the taxonomy.

Linguistic representations are based on existing syntactic-semantic-pragmatic systems of meaning. Linguistic representations, such as speech and text, can, somehow, represent anything and one might therefore wonder why we need any other kind of modality for representing information in HHSI. The basic reason appears to be that linguistic representations lack the *specificity* that characterises analogue representations (Stenning and Oberlander 1991, Bernsen 1995). Instead, linguistic representations are *abstract* and *focused:* they focus, at some level of abstraction, on the subject matter to be communicated without providing its specifics.

Analogue representations, such as images and diagrams, represent aspects of similarity between the representation and what it represents. These aspects can be many, as in holograms, or few, as in a standard data graphics pie graph (or pie chart). Note that the sense of "analogue" in Modality Theory is only remotely related to that of "analogue (vs. digital)". Being complementary to linguistic modalities, analogue representations (sometimes called "iconic" or "isomorphic' representations) have the virtue of specificity but lack abstract focus, whether they be static or dynamic, graphic, acoustic or haptic.

The distinction between *non-arbitrary* and *arbitrary representations* marks the difference between representations that, in order to perform their representational function, rely on an already existing system of meaning and representations that do not.

Static representations and *dynamic representations* are mutually exclusive. However, the notion of static representation used in Modality Theory is not a purely physical one (what does not change or move relative to some frame of reference) nor is it a purely perceptual one (what does not appear to humans to change or move). Rather, static representations are representations that offer the user *freedom of* perceptual

inspection. This means that static representations may be decoded by users in any order desired and as long as desired. Dynamic representations are transient and do not afford freedom of perceptual inspection.

Modality Theory in general and the taxonomy of unimodal output modalities in particular serve the clear and efficient presentation and exchange of information.

The hypothesis that has been confirmed up to this point in the development of Modality Theory and which is inherent to the atomic level of the taxonomy of unimodal output modalities, is a rather strong one. It is that the atomic level fulfils the requirements of completeness, uniqueness, relevance, and intuitiveness stated above.

Any multimodal output representation can be characterised exhaustively as consisting of a combination of atomic-level modalities.

Beyond Literal Meaning. Metaphor and Metonymy

Sometimes it may be preferable to use *non-literal meaning* instead, or in addition, i.e., to use modalities intending them to be understood in a way which is different from their literal meaning. *Metaphoric use* of modalities is probably the best known kind of non-literal use in interaction design so far, such as in the static graphic desktop metaphor. In *metonymy,* a complex subject matter is referred to through some simple part of it.

In general, Modality Theory views non-literal meaning as being derived from literal meaning through subtraction of a smaller or greater amount of the literal connotations of an expression of information in some modality. Modalities can be organised into *modality structures,* such as lists and tables, and modalities can assume *modality roles,* such as when being used as icons. We have seen that modalities can have *non-literal uses* in addition to,

or in replacement of, their literal uses. These phenomena are not exclusive. It is perfectly possible, for instance, to make a table of metaphorical icons.

Modality documents define, explain, analyse, and illustrate unimodal modalities from the point of view of interaction design support. The shared document structure includes the following entries:

(1) Modality profile
(2) Inherited declarative and functional properties
(3) Specific declarative and functional properties
(4) Combinatorial analysis
(5) Relevant operations
(6) Identified types -of
(7) Illustrations

Following the research agenda of Modality Theory, we should at this point address the issue of how to combine unimodal output modalities, unimodal input modalities, and unimodal input/output modalities into usable multimodal representations. However, as the taxonomy of unimodal input modalities is not quite ready yet, this issue will be postponed to the final section of the present chapter.

10.8.1. MULTIMODALITY

Getting a theoretical handle on multimodality would constitute a major result of Modality Theory. As this is work in progress, we are not yet able to present any well tested approach that could be claimed to be superior to, or a valuable complement to, the best current approach.

Modality Theory–based Approaches

How might Modality Theory do better than the best current approach? The theory is superior to the empirical approach in that Modality Theory allows complete *generation* of all possible input, output, and input/output modality combinations at any level,

such as the atomic level, and cross-level as well. However, whilst complete generation is possible in a way that is sufficient for all practical interaction design purposes, the combinatorial explosion involved makes it practically impossible to *investigate* all the generated modality combinations systematically. For instance, if we wanted to investigate all possible n-modal atomic input/output modality combinations where n=10, the number of combinations to be investigated would run into millions. Still, there do seem to be interesting opportunities for exploring this generative/analytic approach by carving out *combinatorial segments* from the taxonomy for systematic analysis, such as a speech-*cum*-other-modalities segment, or an input-manipulation-*cum*-other-modalities segment. These exercises could be further facilitated by tentatively clustering families of similar modalities and treating these as a single modality whose interrelations with other modalities are being investigated. An example could be to treat all analogue static graphics modalities as a single modality, given the fact that these modalities have a series of important properties in common. It is perfectly legitimate to ask questions, such as "How does this particular family of tri-modal combinations combine with other modalities?"

An alternative to the generative approach just described could be to *scale up the SMALTO tool* to address all possible modality combinations. The problem, however, is that this would be likely to produce lists of hundreds of relevant modality properties, creating a space of information too complex for practical use. Part of the usefulness of SMALTO lies in the fact that SMALTO operates with such a small number of modality properties that it is humanly possible to achieve a certain familiarity with all of them, including the broad implications for interaction design of each them, rather quickly. It might be preferable, therefore, to use the SMALTO approach in a slightly different way, i.e. by producing modality properties for limited segments of combinatorial input/output modality space according to current needs, just like SMALTO itself does.

We are currently working on a third approach, which is to *turn Modality Theory as a whole into a hypertext/hypermedia tool* using a common format for modality representation similar to the format described in Section 3 but with an added entry for modality properties. By definition, the tool would include all identified modality properties. The challenge is to make the tool useful for interaction designers who are not, and do not want to become, experts in the theory, for instance by including a comprehensive examples database. In itself, this tool would not constitute a full scientific handle on multimodality in the sense of a systematic approach to multimodal combinations. However, building the Modality Theory tool does seem to be a necessary next step and one that would also facilitate achieving the ultimate goal of mastering the huge space of multimodal combinations.

Finally, a fourth approach is to *analyse the "good compounds"* (Section 6.1) in terms of modality properties in order to explore whether any interesting scientific generalisations might appear.

To complete the research agenda of Modality Theory, we need a well-tested taxonomy of unimodal input modalities, a Modality Theory hypertext/hypermedia tool, and exploration of additional ways in which the theory can be of help in achieving a systematic, creative, and predictive understanding of input/output modality combinations, including those that have not yet been widely used, if at all. These themes are topics of the current research work.

10.9. UsiXML

USer Interface eXtensible Markup Language (UsiXML) is a User Interface Description Language issued from the Cameleon project to specify context-aware, multi-platform, multi-modal user interfaces

Currently, developing the User Interface (UI) of interactive applications is very difficult because of the complexity and the diversity of existing development environments and the high

amount of programming skills required by the developer to reach a usable UI, *i.e.*, markup languages (e.g., HTML), programming languages (e.g., C++ or Java), development skills for communication, and skills for usability engineering.

These difficulties are exacerbated when the same UI should be developed for multiple contexts of use such as multiple categories of users (e.g., having different preferences, speaking different native languages, potentially suffering from disabilities), different computing platforms (e.g., a mobile phone, a Pocket PC, an interactive kiosk, a laptop, a wall screen), and various working environments (e.g., stationary, mobile).

Although designers and programmers are involved in these types of project, the available tools mainly target the developer. Therefore, it is rather difficult for a designer to design a UI for multiple contexts of use while avoiding reproducing multiple UIs for multiple contexts of use. This work proposes a way to separate responsibilities in these types of projects.

UsiXML (which stands for USer Interface eXtensible Markup Language) is an XML-compliant markup language that describes the UI for multiple contexts of use such as Character User Interfaces (CUIs), Graphical User Interfaces (GUIs), Auditory User Interfaces, and Multimodal User Interfaces.

In other words, interactive applications with different types of interaction techniques, modalities of use, and computing platforms can be described in a way that preserves the design independently from peculiar characteristics of physical computing platforms.

UsiXML is intended for non-developers, such as analysts, specifiers, designers, human factors experts, project leaders, novice programmers, and so on.

Of course, UsiXML can also be used by experienced developers.

Thanks to UsiXML, non-developers can shape the UI of any new interactive application by specifying or describing it in UsiXML, without requiring the programming skills usually found in markup

languages (e.g., HTML) and programming languages (e.g., Java or C++).

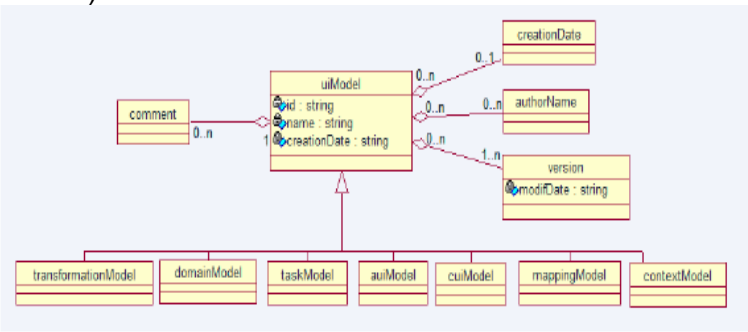


Figure 67: UsiXML Model collection

UsiXML consists of a User Interface Description Language (UIDL), that is to say a declarative language capturing the essence of what a UI is or should be independently of physical characteristics.

UsiXML describes at a high level of abstraction the elements constituting the UI of an application: widgets, controls, containers, modalities, interaction techniques, etc..

UsiXML allows cross-toolkit development of interactive application.

A UI of any UsiXML-compliant application runs in all toolkits that implement it: compilers and interpreters.

UsiXML supports device independence: A: UI can be described in a way that remains autonomous with respect to the devices used in the interactions such as mouse, screen, keyboard, voice recognition system, etc.

If needed, a reference to a particular device can be incorporated.

UsiXML supports platform independence: A UI can be described in a way that remains autonomous with respect to the various computing platforms, such as a mobile phone, Pocket PC, Tablet PC, laptop, desktop, etc.

If needed, a reference to a particular computing platform can be incorporated.

UsiXML supports modality independence: a: a UI can be described in a way that remains independent of any interaction modality such as graphical interaction, vocal interaction, 3D interaction, or haptics.

If needed, a reference to a particular modality can be incorporated.

UsiXML allows re-use of elements previously described in anterior UIs to compose a UI in new applications.

10.10. AMODEUS model for dialogue controller

PAC-Amodeus is a conceptual model useful for devising architectures driven by user-centred properties, including multithreading and multimodality. PAC-Amodeus blends the principles of both Arch and PAC. Arch and its companion, the "slinky" metamodel, provide the appropriate hooks for performing engineering trade-offs such as identifying the appropriate level of abstraction for portability, making semantic repair, or distributing semantics across the components of the architecture. In particular, the five component structure of Arch includes two adapters, the Interface with the Functional Core and the Presentation Techniques Component, that allow the software designer to insulate the key element of the user interface (i.e., the Dialogue Controller) from the variations of the functional core and of the implementation tools (e.g., the X window environment).

The Arch model, however, does not provide any guidance about the decomposition of the Dialogue Controller, nor does it indicate how salient features in new interaction techniques (such as parallelism, fusion, and fission of information) can be supported within the architecture. PAC, on the other hand, stresses the recursive decomposition of the user interface in terms of agents,

but does not pay attention to engineering issues. PAC-Amodeus combines the best of the two worlds.

Succinctly, the five components of the Arch defines the levels of abstraction appropriate for performing engineering trade-offs such as setting the boundaries between the levels of abstraction. We offer the notions of physical device and interaction language as criteria for setting these boundaries. For example, the designer may decide that the Low Level Interaction Component is device dependent. At a higher level of abstraction, the Presentation Techniques Component is device independent but language dependent. At the top of the Arch, the Dialogue Controller is both language and device independent.

PAC-Amodeus refines the Dialogue Controller into a set of co-operative agents that capture parallelism and information processing (e.g., data fusion) at multiple levels of abstraction. In turn, an agent is modelled as a three facet structure:

1. o the Presentation facet is in direct contact with the Presentation Techniques Component of the Arch. It can be used to implement extensions on top of the Presentation Techniques Component;

2. o the Abstraction facet is in direct contact with the Interface with the Functional Core;

3. o the Control facet manages the links and constraints between its two surrounding facets (*i.e.*, the Presentation and the Abstraction facets) as well as its relationships with other agents. As in ALV [8], the Control facet provides the hook for expressing constraints between different perspectives of the same concept.

In combining the Arch principles with PAC, one obtains an "engineerable" model that supports properties inherited from the agent paradigm. Figure 1b illustrates the application of PAC-Amodeus to the software design of MATIS. The Functional Core hosts the database of American cities, airline companies, flight numbers, departure and arrival times, etc. SQL requests are required to access information stored in the database. The

Interface with the Functional Core (IFC) operates as a translator between the SQL formalism and the data structures used in the Dialogue Controller. In MATIS, the IFC serves as a communication bridge. It can also be used to restructure conceptual objects in a form suitable for the purpose of the interaction.

The Dialogue Controller (DC) is organised as a two-level hierarchy of agents. This hierarchy has been devised using heuristic rules. For example, because requests can be elaborated in an interleaved way, there is one agent per pending request.

At the other end of the spectrum, the Low Level Interaction Component (LLIC) is instantiated as two components inherited from the underlying platform: (1) The NeXTSTEP event handler and graphics machine, and (2) the Sphinx speech recogniser, which produces character strings for recognised spoken utterances. Mouse-key events, graphics primitives, and Sphinx character strings are the interaction objects exchanged with the Presentation Techniques Component (PTC).

In turn, the Presentation Techniques Component (PTC) is split into two main parts: the graphics objects (used for both input and output) and the NL parser (used for input only). Graphics objects result from the code generation performed by Interface Builder. The Sphinx parser analyses strings received from the LLIC using a grammar that defines the NL interaction language. As discussed above, the PTC is no longer dependent on devices, but processes information using knowledge about interaction languages.

10.11. References

Reliability ordered cascade

E. Erzin, Y. Yemez, and A. M. Tekalp, "Multimodal speaker identification using an adaptive classifier cascade based on modality reliability," to appear in IEEE Trans. on Multimedia.

Multimodal emotion detection (prosody + facial)

P. Ekman, editor. Emotion in the Human Face. Cambridge University Press, New York, NY, 2nd edition, 1982.

N. Sebe, I. Cohen, T.S. Huang, 'Multimodal Emotion Recognition', in Handbook of Pattern Recognition and Computer Vision, World Scientific, ISBN 981-256-105-6, January 2005.

Mutual information in multimodal SP

T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc, 1991.

Hannes Kruppa and Bernt Schiele. Hierarchical combination of object models using mutual information. In *Proceedings of the 12th British Machine Vision Conference*, pages 103-112, 2001.

D. Roy and A. Pentland. Learning words from natural audio-visual input. In *International Conference of Spoken Language Processing*, December 1998.

Butz Torsten, Thiran Jean-Philippe, 'From error probability to information theoretic (multi-modal) signal processing, Signal Processing, Vol. 85, No 5, pp. 875-902, 2005.

Multimodal biometrics

Czyz Jacek, 'Decision fusion in identity verification using facial images,' PhD Thesis, UCL/TELE, December 12, 2003, 2003.

Czyz Jacek, Bengio Samy, Marcel Christine, Vandendorpe Luc, 'Scalability analysis of audio-visual person authentication,' AVBPA 2003 Audio- and Video-based Person Biometric Authentication Conference, Guildford, UK, June 2003 , J. Kittler, June 9-11 2003, pp. 752-760, 2003.

Multimodal medical image registration

du Bois d´Aische Aloys, De Craene Mathieu, Grégoire Vincent, Macq Benoit, Warfield Simon, 'Efficient Multi-modal dense field non-rigid registration: alignment of histological and section images,' Medical Image Analysis, ITK special issue, 2005

ICARE

Coutaz, J.I., Nigay, L., Salber, D., Blandford, A., May, J., and Young, R.M., Four Easy Pieces for Assessing the Usability

of Multimodal Interaction: t: the CARE Properties, in Proceedings of INTERACT'95 Lillehammer, Norway, 1995.

J. Bouchet, B. Mansoux and L. Nigay, A Component-based Approach: ICARE, ACM-CHI'2005 Workshop on The Future of User, Interface Design Tools, Portland, USA, April 2005.

J. Bouchet, L. Nigay, ICARE: A: a Component-Based Approach for the Design and Development of Multimodal Interfaces, Extended Abstracts of CHI'04, Vienna, Austria, April 2004. pp. 1325-1328

S. Dupuy-chessa, L. du Bousquet, J. Bouchet, Y. Ledru, Test of the ICARE platform fusion mecha-nism, 12th International Workshop on Design, Specification and Verification of Interactive Systems, LNCS, Springer, July 2005.

CAMELEON

Website: http://giove.cnuce.cnr.it/cameleon.html .

Mori, G., Paternò, F., Santoro, C. Tool Support for Designing Nomadic Applications Proceedings of IUI 2003 -Miami, Florida, 12-15 January 2003.

L. Bouillon, J. Vanderdonckt, N. Souchon, Rétro-ingénierie du modèle de présentation pour les pages Web, Revue d'Interaction Homme-Machine, Vol. 3, 2004

Paternò, F. Santoro, C. A Unified Method for Designing Interactive Systems Adaptable to Mobile and Stationary Platforms. Interacting with Computers. Vol. 15, No. 3, pp 347-364, Elsevier, 2003.

G. Mori, F. Paternò, C. Santoro, Design and Development of Multi-Device User Interfaces through Multiple Logical Descriptions, IEEE Transactions on Software Engineering, August 2004, Vol.30, N.8, pp.507-520, IEEE Press.

L. Balme, A. Demeure, N. Barralon, J. Coutaz, G. Calvary, CAMELEON-RT: a: a Software Architecture Reference Model for Distributed, Migratable, and Plastic User Interfaces, EUSAI 2004, LNCS 3295, Markopoulos et al., novembre 2004. pp. 291-302.

Modality theory

Bernsen, N.O. (1994) Foundations of multimodal representations. A taxonomy of representational modalities. Interacting with Computers 6, 347-371

Bernsen, N. O.: Modality Theory: Supporting multimodal interface design. In Proceedings from the ERCIM Workshop on Multimodal Human-Computer Interaction, Nancy, November 1993. ERCIM Workshop Reports 94-W003 1994, 13-23.

UsiXML

Website: http://www.usixml.org

Limbourg Q., Vanderdonckt J., Michotte B., Bouillon L., Florins M., and Trevisan D., USIXML: A User Interface Description Language for Context-Sensitive User Interfaces, in Proceedings of the ACM AVI'2004 Workshop "Developing User Interfaces with XML: A: advances on User Interface Description Languages" (Gallipoli, May 25, 2004), K. Luyten, M. Abrams, Q. Limbourg, J. Vanderdonckt (Eds.), Gallipoli, 2004, pp. 55-62.

AMODEUS

Websitew: http://www.mrc-cbu.cam.ac.uk/amodeus/

Byerley, P.F. Barnard, P.J. and May, J., editors (1993) Computers, Communication and Usability: Design issues, research and methods for integrated services. (North Holland Series in Telecommunication) Elsevier: Amsterdam.

Nigay, L. (1994) Software design and implementation of interactive systems: A case study of multimodal interfaces.Doctoral Thesis, Université Joseph Fourier, Grenoble, France, 29 January 1994.

Paterno, F., editor (1994) Proceedings Eurographics workshop (Design specification and Verifcation of Interactive Systems).

Chapter 11: 10 major practical outcomes

Lausanne Senior Meeting

11.1. Immersive environment

11.1.1. Introduction

In order to achieve the highest level of “naturalness”, users have to be offered “feels like there” sensations and interactivity . Virtual environments, i.e., computer-generated representations of 3D spaces, are probably the best way to reach this objective.

This result can be obtained by the means of hardware worn by the user, such as a head-mounted display and haptic gloves. These devices are nevertheless not very natural, given their multiple wires. Moreover, they are rather fragile and can be used by only one user at a time. It is thus impossible for the people around the user to share his/her experience.

For these reasons, the solution of a small single person immersive display is not very popular, especially for general public applications such as edutainment platforms for museums, expositions, etc. The development of large scale-multiple viewer displays has thus been privileged. As these virtual environments really surround users, they are called immersive environments.

11.1.2. The CAVE

One of the first immersive environments is the CAVE, which was developed by the University of Illinois Chicago and is sold commercially through Pyramid Systems. It is an 8 by 8 by 9 foot room in which the walls and floor are back-projected stereo displays. The user can interact with the virtual environment physically surrounding him by moving, or by pointing with a device called the wand.

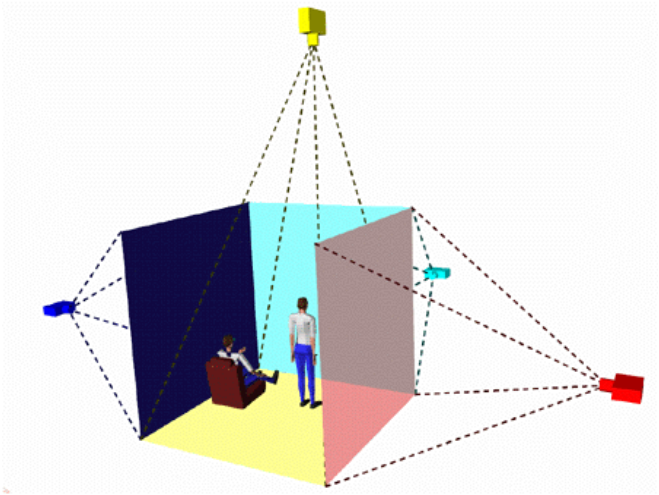


Figure 68: T: the CAVE

11.1.3. Other immersive platforms

Many other platforms have been developed along the CAVE's lines. Some of them differ by the room size or by the number of back-projected walls, others by the sensors used for interaction between user and environment, etc.

Progress is currently being made in the direction of increasing the platform's "naturalness" in its exchanges with users but also by allowing several users to interact with the platform at the same time.

Figure 69: Realization of an immersive environment

11.2. Airplane cockpit

A research project named INTUITION (multimodal interaction integrating innovating technologies) was launched in 2003, for a three-year duration, with funding from the French General Direction of Armament. It brings together three French labs and an industrial partner, THALES-Avionics.

The overall objective of INTUITION is the development of a generic platform to adapt the new technologies in Human-Machine interfaces.

The situation, i.e., the cockpit of a fighter aircraft (namely the RAFALE), carries specific strict constraints such as being usable in real time, since any delay could have dramatic consequences.

Several devices have been envisioned, *i.e.*, a Helmet Mounted Visor, Large Reconfigurable Screen, and Helmet Audio System. Their respective use depends on context information such as the pilot's head position.

However, research is progressing well. The visualisation of information in an aircraft cockpit to assist the pilot in his/her tasks, by allowing quicker reactions, will certainly be more and more widely used, in parallel with specific training.

11.3. Car application

Figure 70: Destination input in a GPS

Nowadays, it is not uncommon to see a car equipped with GPS, cruise control, and DVD reader, and much more room remains for further enhancements! The car is obviously a wide market, but also a very specific environment, where multimodal interfaces can be very useful if correctly designed.

Safety constraints, namely keeping the driver's attention focused on the road, must be very carefully addressed. Other elements also have to be taken into account, such as limitations regarding the possible modalities: the noise level is higher than in a "normal environment" and so speech recognition is more difficult, the driver's position is more or less fixed, which excludes gesturing, and at least one hand is fully occupied by the driving, which imposes maximising contactless interaction.

On the other hand, many applications are possible for multimodal interfaces, from cell phone use to navigation and/or driving assistance, with edutainment for the passengers added on.

A typical example of a multimodal application being developed is the Ford Model U SUV. A multimodal interface centred on speech technology allows drivers to control navigation, make phone calls, operate entertainment features such as the radio or an MP3 player, adjust the climate control and retractable roof, and personalise preferences.

Another one is the new Volkswagen Multimodal Owner's Manual, which received industry-wide accolades at the world's largest automobile trade fair, AutoMechanika (13 - 19 September 2004), in Frankfurt, winning one of the four coveted Innovation Prizes at the fair.

This interface allows the driver to ask questions about the working of his/her car orally while driving. The requested information is then delivered both acoustically and visually through the driver's own mobile device, e.g., a smart phone or mobile digital assistant (MDA), directly from the existing Volkswagen Owner's Manual data located on a content server. Pictures from the manual are delivered simultaneously as a visual aid to accompany the description that is read aloud to the driver. In this way, a passenger, or the driver when the car has been stopped, can compare the picture with the vehicle to locate necessary functional elements more quickly.

11.4. VTK-ITK multimodal medical imaging platforms

11.4.1. The Visualization ToolKit (VTK)

VTK is an open source, freely available software system for 3D computer graphics, image processing, and visualisation. VTK includes a textbook published by Kitware (The Visualization Toolkit, An Object-Oriented Approach To 3D Graphics, 3rd edition ISBN 1-930934-07-6), a C++ class library, and several

interpreted interface layers including Tcl/Tk, Java, and Python. VTK has been implemented on nearly every Unix-based platform, PC's (Windows 95/98/NT/2000/XP) and Mac OSX Jaguar and later. The design and implementation of the library has been strongly influenced by object-oriented principles.

The graphics model in VTK is at a higher level of abstraction than rendering libraries like OpenGL or PEX. This means it is much easier to create useful graphics and visualisation applications. In VTK applications can be written directly in C++, Tcl, Java, or Python. In fact, using the interpreted languages Tcl or Python with Tk, and even Java with its GUI class libraries, it is possible to build useful applications really, really fast.

Finally, the software is a true visualisation system; it doesn't just let you visualise geometry. VTK supports a wide variety of visualisation algorithms, including scalar, vector, tensor, texture, and volumetric methods; and advanced modelling techniques such as implicit modelling, polygon reduction, mesh smoothing, cutting, contouring, and Delaunay triangulation. Moreover, we have directly integrated dozens of imaging algorithms into the system so you can mix 2D imaging/3D graphics algorithms and data.

Our goal is to make the software easy enough for any computer-literate person to use. This includes students, academics, software developers, data analysts, hobbyists, graphics and visualisation users/researchers, engineers, scientists, and researchers. And you have a choice: if you hate C++, then you can use Tcl, Python, or Java.

11.4.2. The Segmentation & Registration ToolKit (ITK)

ITK is an open-source software toolkit for performing registration and segmentation. Segmentation is the process of identifying and classifying data found in a digitally sampled representation. Typically the sampled representation is an image acquired from such medical instrumentation as CT or MRI scanners. Regis-

tration is the task of aligning or developing correspondences between data. For example, in the medical environment, a CT scan may be aligned with a MRI scan in order to combine the information contained in both.

ITK is implemented in C++. ITK is cross-platform, using the CMake build environment to manage the compilation process. In addition, an automated wrapping process generates interfaces between C++ and interpreted programming languages such as Tcl, Java, and Python (using CableSwig). This enables developers to create software using a variety of programming languages. ITK's C++ implementation style is referred to as generic programming (i.e., using templated code). Such C++ templating means that the code is highly efficient, and that many software problems are discovered at compile-time, rather than at run-time during programme execution.

Because ITK is an open-source project, developers from around the world can use, debug, maintain, and extend the software. ITK uses a model of software development referred to as extreme programming. Extreme programming collapses the usual software creation methodology into a simultaneous and iterative process of design-implement-test-release. The key features of extreme programming are communication and testing. Communication among the members of the ITK community is what helps manage the rapid evolution of the software. Testing is what keeps the software stable. In ITK, an extensive testing process (using Dart) is in place that measures the quality on a daily basis. The ITK Testing Dashboard is posted continuously, reflecting the quality of the software.

11.4.3. Multimodal platforms

Several 3D multimodal medical imaging platforms have been and are still being developed using VTK/ITK. These platforms combine images acquired through radiography, PET, US, CT, etc. Here are some of them:

- Medical Studio,

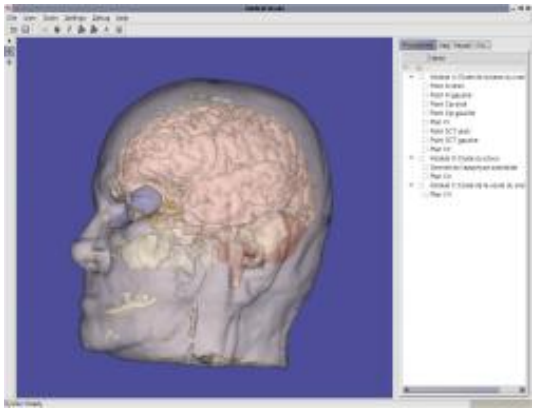


Figure 71

- - Julius,

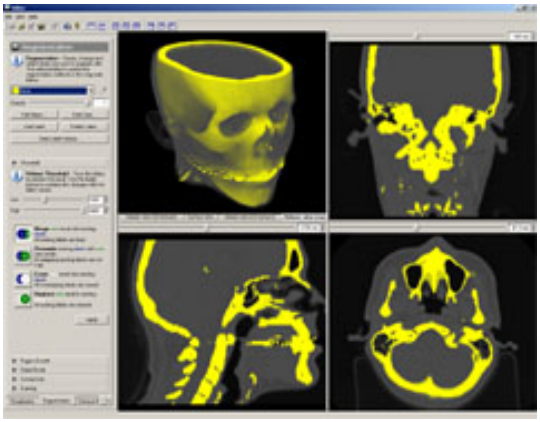


Figure 72

- - OsiriX,

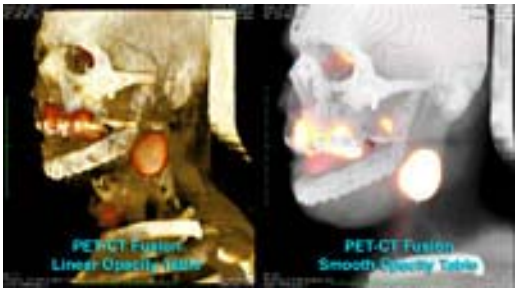


Figure 73

- - MITK,

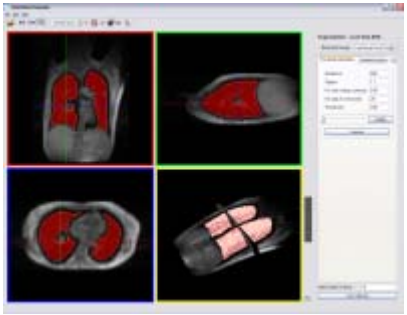


Figure 74

- - CustuX.

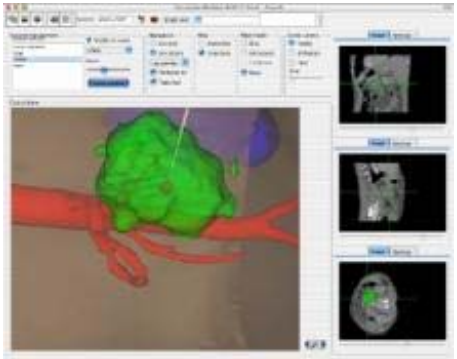


Figure 75

11.5. Augmented reality in surgery

11.5.1. Augmented reality

One of the recent design goals in Human Computer Interaction has been to extend the sensory motor capabilities of computer systems to combine the real and the virtual in order to assist the user in his environment. Such systems are called Augmented Reality (AR).

11.5.2. AR and surgery

Augmented Reality (AR) and Image Guided Surgery (IGS) systems have received considerable attention in many R&D projects in recent years. With IGS systems one can navigate through complex surgical procedures visually with great precision by overlapping on an image of the patient a colour coded preoperative plan specifying details such as the locations of incisions, areas to be avoided, and diseased tissue. This is a typical application of AR systems. The virtual world corresponds to the pre-operative information, the real world corresponds to the intra-operative information, and both should be correctly aligned in real time.

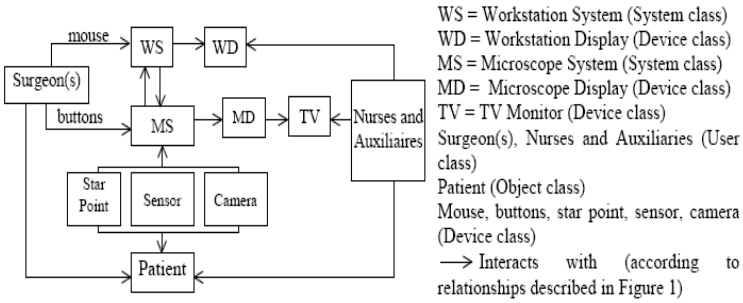


Figure 2. Entities involved in the domain of Image Guided Surgery System.

Figure 76

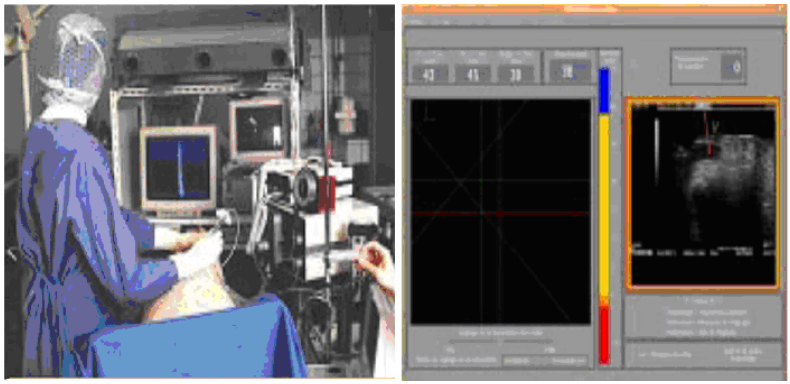


Figure 77. AR used for puncture

11.6. Multimodal mobile augmented reality

11.6.1. Definition

The first AR systems were designed for a specific use in a fixed environment such as the digital desk. The progress made in wireless networks (RF, Radio Frequency and IR, InfraRed, signals) in terms of quality of services now makes it possible to build mobile augmented reality systems.

We believe that mobile AR has a crucial role to play for mobile workers, bringing computer capabilities into the reality of the different workplaces. Let's envision an augmented reality system that will help in deciding where to dig in the streets to access gas pipes. Similar systems already exist.

The user obtains contextual information about the surrounding objects or about a predefined path to follow as he/she walks in a building such as a museum, in the streets, or on a campus,.

Definition: A mobile AR system is one in which augmentation occurs through available knowledge of where the user is (the user's location and therefore the surrounding environment).

Even though the user's location has an impact on the augmentation provided by the system, the latter does not necessarily maintain this location. Indeed, on the one hand, the user's location and orientation are generally known by outdoor systems such as the touring machine system, the position being tracked by a GPS.

On the other hand, for indoor AR systems, objects and places identify themselves to the system (RF, IR, or video-based tags); hence the system does not maintain the user's location. Going one step farther, objects are not only tagged for identification but also contain a mobile code that for example describes the virtual object, i.e., augmentation of the real object.

11.6.2. Applications

Mobile augmented reality systems using multimodal interfaces have already been developed for touring systems, but also archaeology and other collaborative purposes.

The generic concepts used in mobile augmented reality can be applied in other domains such as games. For example, a group of players can search for objects that have been placed virtually in the real environment by other users.

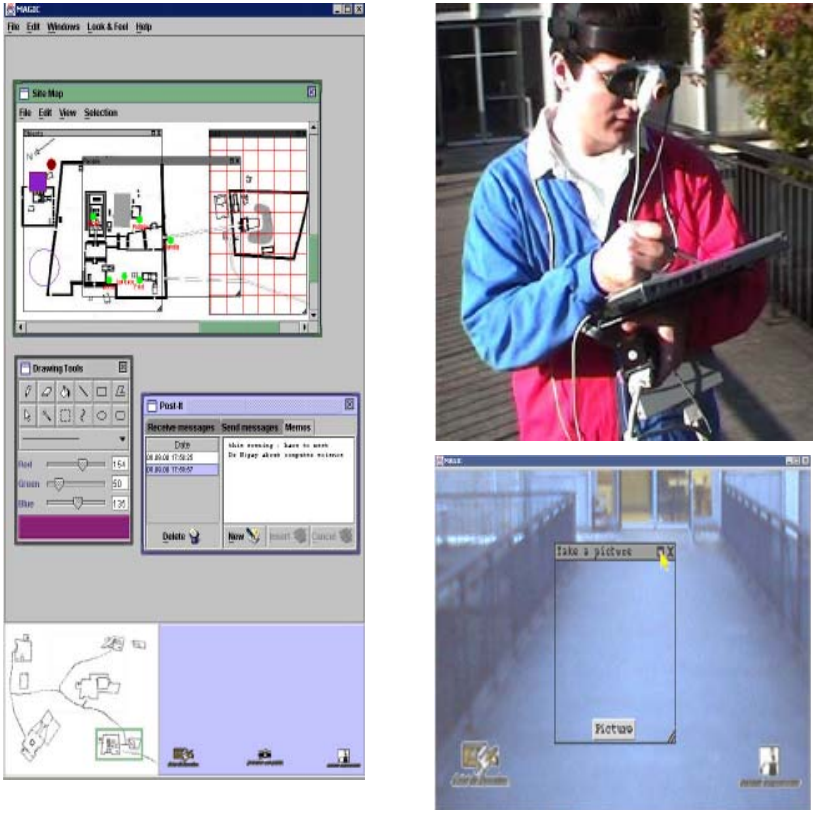


Figure 78: Pictures of MAGIC, a mobile augmented reality system used in archaeology,

11.7. NICE demo

11.7.1. The NICE Project

NICE is aimed at demonstrating universal natural interactive access, in particular for children and adolescents, by developing natural, fun, and experientially rich communication between humans and embodied historical and literary characters. The communication consists of domain-oriented spoken conversation combined with 2D input gestures into a 3D dynamic graphics virtual world inhabited by the fairytale writer Hans Christian Andersen and animated characters from his fairytale world. For the first time, professional computer games technologies have been joined with advanced spoken interaction, and speech recognition technology has been specially developed for recognising the speech and spoken linguistic behaviour of children and adolescents.

11.7.2. The Multimodal Scenario

The NICE PC-based system is designed to enable users to combine their speech and 2D gestures when interacting with characters in an educative game context. It addresses the following scenario: 3D animated lifelike fairytale author Hans Christian Andersen (HCA) is in his 19th century study surrounded by artefacts. At the back of the study is a door that is slightly ajar. This door leads out into the fairytale game world. This world is populated by some of his fairytale characters and their entourage, including, among others, the Naked Emperor and the Snow Queen. When someone talks to HCA, this user becomes an avatar that walks into HCA's study. In the study, the user can have spoken conversations with HCA, including the use of gesture input for indicating artefacts during the conversation. At some point, the user may wish to visit the fairytale world and is invited by HCA to go through the door at the back of the study. Once in the fairytale world, the user may engage in spoken computer games with the characters populating that world, again using 2D gestures as well. The intended users are primarily children and, secondarily, everyone else. The primary scenario of use is in technology and other museums in which, expectedly,

the duration of individual conversations will be 5-30 minutes. Secondarily, we are investigating the feasibility of prototyping the world's first spoken computer game for home use with its average of 30 hours of user interaction time.

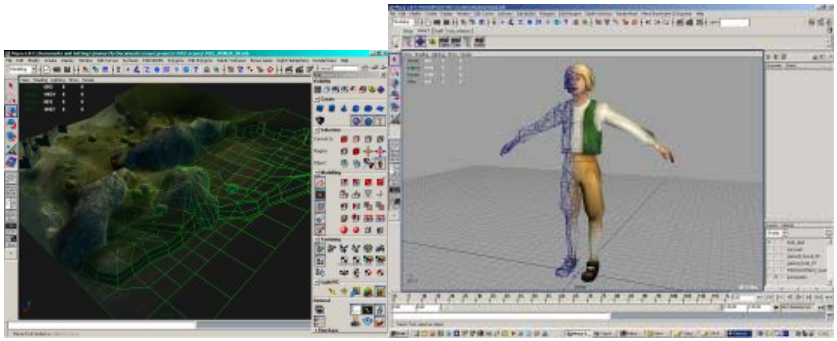


Figure 79: Screenshots of NICE demo

The primary research challenge taken up by NICE is to move from the existing paradigm of task-oriented spoken dialogue with computer system to the next step, which we call domain-oriented spoken dialogue. In domain-oriented spoken dialogue, there is no user task any more to constrain the dialogue and help enormously in its design and implementation, but only the semi-open domain(s) of discourse which, in the case of HCA, are his life, fairy tales, 3D physical presence, modelling of the user, and role as kind of gate-keeper for the virtual fairytale world.

In a limited fashion, however, we are also investigating the use of combined input speech and 2D gestures for indicating objects and other entities of interest.

11.8. FAME: context-aware distributed user interfaces

11.8.1. What is it?

FAME (Facilitating Agents in Multicultural Exchange) is a joint development of IIHM and Prima laboratories (France), in the context of the European Project FAME IST-2000-28323 (http://isl.ira.ika.de/fame).

FAME is an example of interactive space compliant with the principles of ambient intelligence: the interaction with the computer is no longer confined to mouse-keyboard, but comes within the context of a whole environment (tables, walls, and other familiar objects) to create an ambiance. In this environment the digital and real worlds work in synergy to serve us better. The FAME interactive space has ears and eyes (microphones and cameras) to observe us, predict our actions, and act in a relevant way. FAME allows the discovery, while having fun, of information spaces, in this case the major research axes of IMAG.

11.8.2. How does it work?

Figure 80

On the ceiling, a video projector displays a graphical representation of the information space on a table[5]. A camera follows the tokens' handling; the tokens can be placed on the table to select subjects of interest. The details concerning the selected subject are projected on a wall.

When people are conversing, FAME is able to identify their discussion topics and displays them as a bubble fountain on the wall. If a subject is deemed interesting by someone, it can be selected for more information. Then FAME projects complementary information on another wall by searching the Internet or an information base.

It is also possible to take part in the dynamic construction of the information space by asking FAME's Smart Cameraman to record one's testimony. These recordings are made according to the rules of cinematography. They are then replayed on demand. Behind the scene, FAME requires 4 top class computers (PC and MacIntosh) linked via a LAN, 3 video projectors, a camera, and a microphone.

11.8.3. What is it used for?

The research is aimed at concept identification, software technique creation, and setting up the necessary tools and methods for creating ambient interactive spaces.

An ambient interactive environment is a space with perception, action, and communication skills in order to offer services adapted for human activities. It is characterised by:

1) the computer's disappearance, so that its utilisation is more natural for the human being,

2) the possibility for the user to build his/her space opportunistically by resources composition and reconfiguration, given that the space is now limited to a workstation,

---

[5] representation of the information space as a set of tokens and a table (??) – il faut introduire les "tokens" avant de partler de leur manipulation!

3) the provision of services adapted to the interaction context, as the system is currently unable to adapt to the situation.

Figure 81

These general objectives are dependent on advances in

- artificial perception, so that the system is able to localise the user, to locate objects in space, and to track a finger in real –time, regardless of the light conditions;
- interaction techniques, with study of the balance between explicit and implicit interaction; and
- software models and techniques allowing the adaptation of all system components. The notions of context, machine learning, and flexibility are very important here.

11.9. ConcurTaskTrees

11.9.1. What is it?

ConcurTaskTrees is a notation for task model specifications that has been developed by Fabio Paternò (IST, Italy) to overcome limitations of notations previously used to design interactive

applications. Its main purpose is to be an easy-to-use notation that can support the design of real industrial applications, which usually means applications with medium-large dimensions.

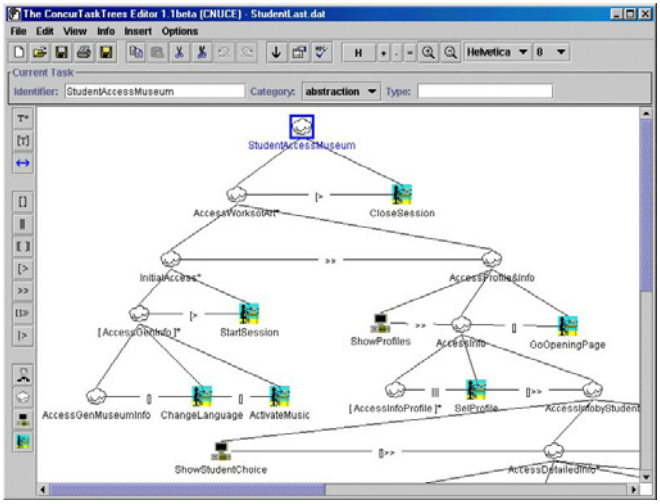


Figure 82: ConcurTaskTrees editor

The main features of ConcurTaskTrees are:

- *Hierarchical structure*. A hierarchical structure is something very intuitive. In fact, when people have to solve a problem they often tend to break it down it into smaller problems while maintaining the relationships among the smaller parts of the solution. The hierarchical structure of this specification has two advantages: it provides a large range of granularity, allowing large and small task structures to be re-used; and it enables re-usable task structures to be defined at both a low and a high semantic level.
- *Graphical syntax.* A graphical syntax often (not always) is more easy to interpret. In this case it should reflect the logical structure, so it should have a tree-like shape.
- *Concurrent notation. O*perators for temporal ordering are used to link subtasks at the same abstraction level. This

sort of aspect is usually implicit, expressed informally in the outputs of task analysis. Making the analyst use these operators is a substantial change to normal practice. The reason for this innovation is that after an informal task analysis we want designers to express clearly the logical temporal relationships. This is because such ordering should be taken into account in the user interface implementation to allow the user to perform at any time the tasks that should be active from a semantic point of view.

- *Focus on activities.* This allows designers to concentrate on the most relevant aspects when designing interactive applications that encompass both user and system-related aspects, avoiding low levels implementation details that at the design stage would only obscure the decisions to take.

This notation has shown two positive results:

- *an expressive and flexible notation* able to represent concurrent and interactive activities, also with the possibility of supporting co-operation among multiple users and possible interruptions.
- *Compact, understandable representation.* The key aspect in the success of a notation is the ability to provide much information in an intuitive way without requiring excessive efforts from the notation's users. ConcurTaskTrees is able to support this, as demonstrated by its use by people working in industry without a background in computer science.

11.9.2. Applications

Environments have been built that support task models specified by the ConcurTaskTrees [5] notation.

Such tools have been developed within the GUITARE R&D European project (http://giove.cnuce.cnr.it/guitare.html), which aims

to improve the design of Enterprise Resource Planning applications. These tools are then applied in other projects related to other application domain areas, for example, in the MEFISTO project (http://giove.cnuce.cnr.it/mefisto.html) concerning Air Traffic Control applications and the Safeguard of Cultural Heritage project (http://giove.cnuce.cnr.it/bc.html) concerning museum applications.

11.10. GrafiXML

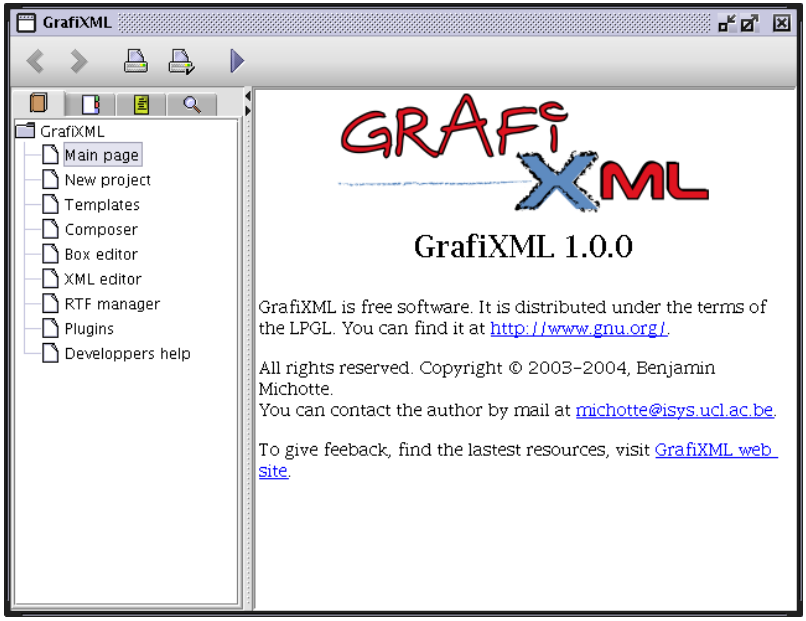


Figure 83: GrafiXML

GrafiXML is a graphical tool developed at UCL (Belgium). It is a user interface editor for multiple platforms, languages, and modalities

As editing a new UI in USIXML directly can be considered a tedious task, a specific editor called *GrafiXML* has been

developed to cope with the development of USIXML models. As USIXML is at first hand a textual language, an *ad hoc* XML editor was created. In this editor, the designer can draw in direct manipulation any graphical UI by directly placing CIOs where they need to be and editing their properties in the Composer, so that they are instantly reflected in the UI design. The designer may want to see the corresponding USIXML specifications and edit them at any time,. Selecting a USIXML tag automatically displays possible values for this tag in a contextual menu. When the tag or elements are modified, this change is propagated in the graphical representation. In this way, bidirectional mapping is maintained between a UI and its USIXML specifications: each time a part is modified, the other one is updated accordingly.

11.11. References

Immersive environment

Cruz-Neira, C., Sandin, D.J., DeFanti, T.A. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. ACM Computer Graphics, Vol. 27, Number 2 , July 1993.pp 135-142.

Cavazza Marc, Martin Olivier, Charles Fred, Mead Steve, Marichal Xavier: 'Multi-modal Acting in Mixed Reality Interactive Storytelling,' IEEE Multimedia Magazine, July-September, 30-39, 2004

Airplane cockpit

J. Bouchet, L. Nigay, T. Ganille, The ICARE Component-Based Approach for Multimodal Input Interaction: Application to realtime military aircraft cockpits, Conference Proceedings of HCI International 2005, the 11th International Conference on Human-Computer Interaction, Las Vegas, Nevada, USA,july 2005. Lawrence Erlbaum Associates. In press.

Car applications

Bernsen, N. O. and Dybkjær, L.: Exploring natural interaction in the car. Paper to appear in Proceedings of the CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation, Verona, Italy, December 2001.

VTK-ITK platforms for 3D multimodal medical imaging

www.medicalstudio.org

http://julius.caesar.de/home/home.html

http://homepage.mac.com/rossetantoine/osirix/Index2.html

http://mbi.dkfz-heidelberg.de/mitk/screenshots.html

Augmented reality in surgical operating theatres

E. Dubois, L. Nigay, J. Troccaz, O. Chavanon, L. Carrat, Design Space for Augmented Surgery, an Augmented Reality Case Study, Proceedings of INTERACT'99, Edinburgh, A. Sasse & C. Johnson Eds, IFIP IOS Press Publ., 1999, pp. 353-359.

Trevisan, D., Vanderdonckt, J., Macq, B., Raftopoulos, C., Modeling Augmented Reality System, Image Guided Surgery Case Study, Proceedings of the 1st Glasgow Context Group Colloquium GCG'2002 «Building Bridges: Interdisciplinary Context-Sensitive Computing» (University of Glasgow, 9 September 2002).

Multimodal mobile augmented reality

P. Renevier, L. Nigay, Mobile Collaborative Augmented Reality: t: the Augmented Stroll, Proceedings of EHCI'01, IFIP WG2.7 (13.2), Working Conference, Toronto, may 2001, Lecture Notes in Computer Science (LNCS), Vol. 2254, Springer-Verlag Publ., pp. 315-334.

P. Renevier, L. Nigay, J. Bouchet, L. Pasqualetti, Generic Interaction Techniques for Mobile Collaborative Mixed Systems, in Computer-Aided Design of User Interfaces IV, Kluwer Academic Publishers, 2005, pp. 307-320.

NICE demo

Website: www.niceproject.com

Bernsen, N. O. and Dybkjær, L.: Structured interview-based evaluation of spoken multimodal conversation with H.C. Andersen. Proceedings of The International Conference for Spoken Language Processing, ICSLP 2004, South Korea, October 2004. International Speech Communication Association (ISCA) 2004, Vol. 1, 277-280.

FAME

Petra Gieselmann, Matthias Denecke: Towards Multimodal Interaction with an Intelligent Room. In: Proceedings of the Eurospeech, Geneva, 2003.

H. Holzapfel, K. Nickel, R. Stiefelhagen, Implementation and Evaluation of a Constraint Based Multimodal Fusion System for Speech and 3D Pointing Gestures, ICMI 2004, State College, PA, USA.

ConcurTaskTree

F.Paterno', C.Mancini, "Developing Adaptable Hypermedia", Proceedings ACM Intelligent User Interfaces, pp. 163-170, Los Angeles, January '99, ACM Press.

GrafiXML

www.usixml.org/index.php?view=page&idpage=10

Chapter 12: 10 major challenges

Lausanne Senior Meeting

12.1. Social interactions

Some recent studies are dedicated to defining patterns for designing conversational or even interfaces, which are somehow still underdeveloped for the moment, despite the progress in related technologies, such as emotion and speech recognition, and Text-To-Speech interfaces. Until recently, the overall interactive cycle between human and computer was ignored.

The specificity of these interfaces is indeed their high level of interactivity with the user(s). Consequently, in addition to be able to control all input and output technologies, they must be as adaptive and reactive as possible, because users' behaviours vary with the individual and circumstances and can adapt to the behaviour of the animated persona that they are facing.

The empirical studies that have been conducted recently are trying to enhance the knowledge related to these changes and, accordingly, to improve the design of multimodal interfaces.

First, people do not automatically use several modalities when given the possibility; they are indeed more likely to interact "unimodally" rather than multimodally when continuing a dialogue that has already been initiated with the system or answering an easy request. The choice among available modalities also

depends on several factors: context (mobile use or not), type of request, and individual differences among users. Moreover, when it comes to acting multimodally, people do not all have the same behaviour. Some people prefer to use all modalities together, synchronously, whereas others prefer to use them sequentially. This small difference can have a major impact on the system's interpretation of the inputs. Finally, user category-specific differences also exist. So, elderly users, for example, will use self-speech when confronted with harder requests, which very concretely may require a different configuration of the speech recognition software.

12.2. Software Architecture reference model

Some software architecture reference models have already been proposed for multimodal applications. The authors are nevertheless of the opinion that these existing reference models do not address all the key issues that they have identified, to wit,

- explicit definition of fusion/fission mechanisms for input and output;
- different levels of abstraction where the fusion/fission takes place: lexical, syntactic, semantic, task fusion (could be based on Nielsen's linguistic model of interaction);
- explicit support for reconfigurability for better context-adaptation (especially at runtime as opposed at design time): self-description, exchange protocol; and
- support for multi-* concerns: multi-user, multi-device, multi-platform, multi-media, multi-modal, and so on.

As of now, no single reference model addresses all these requirements.

Standardisation work in this direction is currently being conducted within the W3C working group on multimodal

interfaces, with the collaboration of several SIMILAR members. It is nevertheless far from complete.

12.3. Mix of architectural styles

Several architecture models exist for multimodal interfaces, but so far mixing them in order to improve their characteristics has not been attempted. The authors think this could be a path to explore.

12.4. Software component description

(cfr. CDIL)

a. Services they provide
b. Services they request
c. Resources requested
d. Quality of Service (QoS): stability, latency, and precision

12.5. Polymorphic user interfaces

Interfaces are in principle designed to fit their target users' requirements, skills, etc. But it is not always feasible to have one single interface per user category, *e.g.*, children, adults, people with disabilities, etc. Furthermore, the context also intervenes in the design of the interface, and total control over the context is not possible. Polymorphic UI have several different independent representations. They are thus highly interesting in considering these aspects.

12.6. Genericness of model, methods, and tools

As of now many tools and methods have been and are still being developed, while different models co-exist. Some of them cover

different areas, but others greatly overlap. This competition offers the advantage of leaving room for choice, but it also slows down development because of the lack of genericness. This issue should thus be addressed.

12.7. Scalability

Currently most of the multimodal interfaces that are developed are limited when it comes to the platform(s), number of users, and/or modalities. This lack of capability for extension obviously limits their potential for large adoption.

Scalability should be considered when developing a multimodal interface, in order to enable and facilitate its upgradability.

12.8. Reference Glossary

To bring HCI and DSP closer to each other, which is highly desirable, not to say necessary, for future research in the area of multimodal interfaces, one of the first things to do is indisputably to set up a common vocabulary.

Until now, these two worlds have indeed been using the same words, but their definitions were not always the same, even within the same discipline. With such different meanings, communication and understanding between the two communities effectively remain difficult.

This issue was highlighted at SIMILAR's Lausanne meeting, where two days of presentations were spent understanding what exactly was covered by the notions of fusion and fission, which are central to multimodal interface development, in each community.

The common vocabulary should ideally cover the most usual areas of Multimodal Applications and User Interfaces.

Such unification work has already been initiated in the CAMELEON Project (Context Aware Modelling for Enabling and

Leveraging Effective interactiON), a Shared-Costs RTD IST Project, which has given rise to a glossary that can be consulted at:

http://giove.cnuce.cnr.it/cameleon/glossary.html

Nevertheless, the task is far from finished:

- CAMELEON is more HCI oriented, meaning that additional input from the DSP community is needed.
- Some terms are still deemed “dangling”, as no unanimous agreement on their definitions could be reached.
- Some terms are still missing.

SIMILAR has thus decided to join CAMELEON in this challenging work.

For this purpose, meetings of senior researchers from HCI and DSP inside the NoE are planned, in constant interaction with the outside world through the SIMILARnet portal, which will reflect the latest state of the definition work, but also through contacts with internationally acknowledged experts in the area of multimodal interfaces.

The hope is that this continuous communication will culminate in the adoption of a set of definitions that is widely accepted both inside and outside the consortium after which this set will gradually be extended to cover all the key words used in multimodal interfaces.

All editions of the SIMILAR Dreams book will also include the latest version of the glossary: this will constitute both a valuable dissemination tool and a simple way to check the work’s progress simply by comparing the glossaries of the different editions.

12.9. Generic contactless multimodal interfaces

Contactless multimodal interfaces are considered a valuable choice, notably for edutainment applications. They effectively avoid all physical contact between the user and the system, which greatly lowers the risk of the equipment's being damaged. Despite this undisputable advantage, to date no generic contactless multimodal interface exists.

12.10. Security/privacy in context-sensitive multimodal applications

In order to interact with users efficiently, multimodal interfaces have to capture a wide range of information concerning them, from their explicit instructions to their location, the emotions they express, etc. As not all this information is consciously emitted by the user, privacy issues may obviously arise. What happens if a user does not want the system to capture what he is talking about, for example? Security is also an important question to be handled, especially in sensitive domains such as military and medical multimodal applications. A multimodal interface should at least make the user aware of what is being captured about him/her, and provide measures to avoid any illegitimate use or processing of this information.

12.11. References

Katashi Nagao and Akikazu Takeuchi. 1994. Social interaction: Multimodal conversation with social agents. In Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAL 9d). The MIT Press.

www.iis.ee.ic.ac.uk/~j.pitt/teaching/HCI_AffectiveComputing_2002.ppt

Oviatt, S.L, Darves, C. & Coulston, R. "Toward Adaptive Conversational Interfaces: Modeling Speech Convergence with Animated Personas," in ACM Transactions on Computer-Human Interaction (TOCHI) 11(3) September 2004.

www.w3.org/TR/2005/WD-mmi-arch-20050422/

H. Djenidi, A. Ramdane-Cherif, C. Tadj, N. Levy. "Dynamic Based Agent Reconfiguration of Multimedia Multimodal Architecture," *mse*, vol. 00, no. , p. 198, IEEE 2002.

Constantine Stephanidis, Demosthenes Akoumianakis, Universal design: towards universal access in the information society, CHI '01 extended abstracts on Human factors in computer systems, March 31-April 05, 2001, Seattle, Washington

http://giove.cnuce.cnr.it/cameleon/glossary.html

Chapter 13: Interviewing

13.1. Questionnaire

The collaboration of several SIMILAR researchers has led to the following set of questions on MMI:

1. Brief description of your domain (including actual needs/requirements)
2. Target group (professionals, public at large, people with disabilities, surgeons, etc.)
3. Actual modalities available in your interfaces, considering that modality = language (text, graphic, sound, 3D, etc.) + device (computer, robot, PDA, head mounted display, etc.)
4. The ideal interface in specific situations: major challenges in your domain. What interactions are envisioned in the near future (new modalities?)
5. How do you see the integration of the following aspects in current/future multimodal interfaces:
    a. Context-aware adaptation
    b. Continuity across media and devices and task continuity
    c. What is the best way to combine modalities in input and output?
    d. How to assess usability?

e. Personal interfaces *versus* group interfaces (collaborative interaction)
f. Task help, task surveillance, task monitoring (for assessment)
g. Trust in the interface (what is the required level of confidence?)
h. Interfaces for learning

All our contributors have received this questionnaire. Their responses will be presented, either as usual questions and answers, or as a text (all in one go) stating their opinions on current and future considerations about MMI. The form of presentation that is chosen will depend of the content of our gurus' answers: some of them did not have something to say on all questions, which is legitimate given their broad spectrum, while transverse issues that we feel deserve to be highlighted emerged from some other respondents' answers.

13.2. Industry

13.2.1. Xavier Marichal (Alterface, Belgium)

Xavier Marichal

BIO

PhD in Electrical Engineering, Xavier Marichal has been the Technical Director (CTO) of Alterface, a Belgian company that he founded, since 2002. The company's activities consist in developing interactive settings (software platforms) for education, entertainment, information, and advertising.

He is also the author of several publications on mixed reality.

Within Alterface, Mr Marichal is developing natural interaction set-ups for the general public. The goal is to entertain and/or educate people in the most spectacular and immersive way, without having to handle dedicated hardware and peripherals.

Figure 84: Two recent achievements of Alterface:
A seismic simulator (Vulcania, France)
& an interface for sorting garbage

Through their proprietary interaction engine, Salto, multimedia scenarios involving many modalities (sill images, videos, 3D scenes, sounds, lighting effects, etc) are fed information acquired through many, possibly interconnected, sensors, in addition to various analysis modules. Currently, new modalities, such as improved gestures, are being developed. Multimodality is thus on the way, for input as well as output.

Context is already instantiated through weather stations, biometric data, etc. and context-aware adaptation is deemed essential.

Group interfaces allowing collaborative interactions are seen as an abstract layer on top of individual interaction, which enables “intelligent” collaboration/co-operation among several users, beside their individual behaviour. They are key for success.

Users’ trust in the interface is not really a problem in a non-critical (security or health) field such as edutainment. In this specific context, the issue consists more of ergonomics and usability: “if people do not catch the way it works in a few seconds, they just leave it with a ‘Pfuh! It does not work’ assertion.”

In Mr Marichal's opinion, the major challenge is total elimination of the interface, so that the user can interact with the system simply by behaving normally.

13.2.2. Frederic Kaplan (Sony, France)

Frederic Kaplan with an entertainment robot.

BIO

Frédéric Kaplan has a PhD in computer science.

Since 1997, he has been working as a researcher at Sony Computer Science Laboratory in Paris, where his work focuses on artificial intelligence and the integration of new machines in society. He has developed new entertainment robot prototypes and is the author of two books , *Les machines apprivoisées*" (Vuibert, 2005) and *La naissance d'une langue chez les robots* (Hermès, 2001).

Dr Kaplan's current work concerns developmental robotics, which means designing robots able to learn from their environment and to become autonomous members of a community. These robots, which are above all dedicated to general public entertainment, are equipped with a wide range of sensors for better perception of their environment. They are also able to interact with other types of computer equipment through remote connections.

The future of developmental robotics involves new types of sensor. For example, artificial whiskers are being developed in order to improve these robots' environmental perception: (for texture recognition, touch, speed, airflow, and waterflow).

Perception is going to become more and more active, which means that robots will be able to detect events by directly observing the effects of their own actions on the environment

and during remote interactions, for example through the Internet. As they interact, such robots will develop through their own “life trajectories”. With this kind of technology, each robot will become unique, its behaviour resulting from the history of its past interactions.

13.2.3. Stéphane Chatty and Stéphane Sire (Intuilab, France)

Stéphane Chatty

Stéphane Chatty is a co-founder and chief technical officer of IntuiLab, a company created in 2002 to make multi-modal, natural, and mobile interaction available to industrial companies. He holds a degree from Ecole Polytechnique (Paris, France) and a PhD in computer science from Université de Paris Sud. From 1992 to 2002 he worked at the French research centre on air traffic control, where he created a research team on user interaction and then directed a research department on tools for en-route air traffic controllers.

At IntuiLab he has been involved both in research projects on software engineering for user interfaces and design projects for car cockpits, process supervision systems, end-user telephony, aeronautics, and defence.

Stéphane Sire is a research engineer in Human-Computer Interaction. He joined IntuiLab in Toulouse, France, in 2003, to develop a new generation of user-interface programming tools and languages for multimodal interaction. He has an industrial and academic background with participation in research projects on Computer Supported Co-operative Work (groupware), e-learning and usability evaluation of e-learning environments, wizard-of-Oz experimentation, multimodal interaction, and electronic document models.

Stéphane Sire

From 2000 to 2003 he was seen at various places in Switzerland and the Netherlands, where he was doing post-doctoral research. He earned a PhD in computer science in 2000 and a computer science engineering degree from the ENSIMAG in Grenoble in 1994. His chief interest now is in developing languages for programming user interfaces.

1. Brief description of our domain

IntuiLab, created in 2002, is a company specialised in innovative user interfaces for complex systems as well as in multimodality and mobility technologies. Its 20 person team is composed of user interface specialists, such as human factor specialists, designers, graphics specialists, programmers, PhDs, and researchers.

It is one of the first companies to provide the methods and tools required to design and develop multimodal and mobile interfaces:

- D (means “challenge”), a participatory design method based on rapid prototyping.
- Intuikit®, a software environment for prototyping and developing multimodal and mobile user interfaces, supporting the D methodology.

Intuilab is strongly involved in R&D activities such as the development of its "Model Based Architecture and Development" for advanced graphics and multimodal MMIs and methodological research to make multimodal user interfaces more intuitive.

2. Target group

Currently, the interfaces designed and developed by IntuiLab target both professional and mass-market users.

For the professional market, the interface should enable rapid, efficient interaction. This is why we address all activities in which the human factor plays a key role, such as supervision systems, decision making systems, and so on.

For the mass-market the interfaces should be intuitive, friendly and easy to use. These interfaces may be embedded on mobile devices (such as PDAs or telephones) or be part of telematic systems (e.g., in cars or homes).

3. Actual modalities available in our interfaces

Today, IntuiLab uses a panel of modalities in its projects. The main modalities currently used are:

- Touch-screen interaction
- Voice recognition and synthesis
- Gesture recognition
- Use of advanced, realistic, and professional graphics (transparency, textures, outlines, shadows, shadings, gradient, lights, etc.)

- Animated visual feedback (trajectories, rhythm, elasticity)

IntuiLab has also developed soft components or styles that use each of these modalities, such as:

- Classic styles, like WIMP components based on pointer manipulation;
- Advanced direct manipulation with gesture recognition; and
- Multimodality

IntuiLab is currently working on a framework for developing reusable multimodal widgets (combining graphics, voice and gesture).

Our user interfaces can run on computers (desktop and embedded systems), PDAs, and Smartphones with different OS (Linux, Windows, MacOS).

4. Ideal interface in specific situations: major challenges in your field. What interactions are envisioned in the near future (new modalities?)

The ideal interface would be a “natural” interface. That is to say a flexible interface that enables the user to choose the most appropriate modality for each specific situation to optimise usability and efficiency.

Nowadays, the “emerging” modalities seem to be voice and pen or finger input on a touch screen. Voice is already used in a few interfaces, such as the address book in mobile phones, but is not yet common. We do not know how fast voice and multimodality are going to be adopted for user interfaces. That will certainly depend on the availability of convincing applications. Voice-enabled mobile phones with tactile input could become the first mass-market device from which users will get accustomed to multimodality.

Future interfaces should offer both modalities (voice and tactile interaction) and also integrate location-based services, but not

be dependent on multimodality. This way users could start experimenting with these new styles.

The near-future interfaces will have to take industrial maturity into account. They should be as natural as possible and their industrial production should be feasible. In this perspective, a great challenge is to bridge the gap between system design and interface design, in order to avoid system specification redundancy. We have started to think of mapping software development engineering cycles on system engineering with more model-based approaches to multimodal user interface design, like they do in web services, but without losing the creativity inherent in HCI projects.

5. How do you see the integration of the following aspects in current/future multimodal interfaces :

a. Context-aware adaptation

Context-aware adaptation is clearly useful, especially to adapt information delivery and request interpretation to location knowledge. But we believe that this kind of interface will not appear suddenly in the near future.

It will be a step-by-step development. The first step will consist of a programmed adaptation rather than an automatic or “intelligent” adaptation. The interface will be able to detect and react to specific situations thanks to a programmed reply, but not take “human” decisions. The most important thing is that the user should always be in control of automatic adaptation, because s/he could rapidly be lost, not understand why a change has occurred, and finally misuse the application.

The first step will also have to go through standard modelling of the context. Indeed, if we consider that the multiple dimensions of the context define a map, then it's important to identify the different territories on that map where a given user interface will remain usable without adaptation. For that reason it is primordial

to make progress on defining a context and its boundaries. Some initiatives to define the so-called device profiles have started at OMA and W3C and we expect they will grow up to include more dimensions in the definition of the context.

b. Continuity across media and devices and task continuity?

With the advances of technology, convergence of media and devices seems to be unavoidable. But what will this “cross-technology” consist of?

The question of task continuity is highly complex. We believe that it is very difficult but essential to give the user the feeling that s/he is using the same service, even if s/he goes from one device to another. The characterisation of two equivalent interfaces is also a question we are interested in.

We also believe that this cross-technology will not culminate in a unique device with all functionalities, but rather in the advent of “bridging devices” that will allow users to carry their “electronic worlds” in their pockets and to converse with any device. This “physical embodiment” of the user's multiple access points in their electronic realms could take the form of a mobile phone, a USB key, an iPod, or - why not? - a finger. The only thing that is important is that this artefact will be used to carry information, the right to access information, or the continuation of a task, from one device to the other, a little like pen, paper, and keys in the real world.

c. What is the best way to combine modalities in input and output?

According to the various applications that we have already developed, there is more evidence that alternate and sequential synergic multimodality is more useful than parallel synergic multimodality (fusion) in the short term. Why is that? Well, firstly because multimodal fusion increases the complexity level of the

interface. Secondly, if you use combined modalities you often end up with user frustration, because it is not always possible to use this combined mode, or modality, depending on the situation s/he is in. For instance, think of dictating a very private message in a train full of passengers. Finally, we have noticed that there is a natural user affinity to focus on a single modality.

Now, it would be interesting for the future applications to be able to switch modality in the course of a long task. For example, you choose to start the task using pen based interaction, and realise that it would certainly be easier to say what you want to do, and switch to a voice interaction.

The output modalities should be available in parallel, in order to be efficient regardless of the user's situation. Parallelism of output modalities is really efficient only if the indications given by each modality are complementary. Otherwise, it only creates redundancy and uselessly increases user workload.

d. How to assess usability?

As with other user interfaces; three usability aspects have to be investigated:

- Learning or ability to learn
- Efficiency
- Satisfaction

With multimodal user interface, learnability is a great challenge today, especially for mass-market applications, since it is often difficult for users to know how to use the available new modalities (voice, gesture recognition, and so on) efficiently. Assessing multimodal user interface efficiency may be based on well-known measures (task success, task time, error rate, etc.) but may also involve new analyses. For instance, it may be useful to study time span variability between voice and gesture when the two modalities are combined in a single act.

Actually, given our current lack of knowledge about user behaviours regarding multimodal user interfaces, we strongly

believe that iterative design is crucial. It allows one to obtain user feedback continuously in order to validate design choices before the development begins. In this perspective, a specific Wizard of Oz platform for multimodal user interface should be developed to enable early usability assessment.

e. Personal interfaces versus group interfaces (collaborative interaction)

We have the feeling that group interfaces already exist in many domains, like in video games, and all the services offered to teenagers. Now, in industrial applications, it is an uncommon emerging tendency that will sooner or later be unavoidable. On the other hand, international working groups and open source communities are already using collaborative tools. They are getting excellent results with very classic user interfaces. This is why we believe that multimodal and innovative interaction styles could be of a great benefit to these groups.

f. Task help, task surveillance, and task monitoring (for assessment)

The best way to assess multimodal interfaces is certainly to do “on site” experiments conducted in workplaces. Thanks to mobile, wireless, and portable equipment such as webcams, it is possible to have a “mobile usability lab”. It is then possible to observe the user, the way s/he uses the interface, and how s/he reacts, in real situation of use and to disturb her/him as little as possible.

g. Trust in interface (what is the required level of confidence?)

For the interface itself and the interaction style, the level of confidence, reliability, and accuracy required depends on the kind of task carried out. For example, a selection task is less critical than a validation task, so the level of confidence for this last task needs to be higher. In fact, it is the designer’s job to

choose carefully which modality should be offered for which part of the dialogue between the user and the machine.

Now, trust in the interface is necessarily linked to trust in the service provider behind the interface. The interface is the reflection of this supplier. If the interface has advanced graphics, seems to be sure, and inspires confidence, the user will automatically trust the service provider (for online business, for example).

h. Interfaces for learning

We believe that multimodal interfaces will reinforce the “being there” feeling in remote training sessions. With a multimodal and natural interface, it will be possible to have access to education anywhere at anytime and have more flexible timetables.

13.2.4. Robin Springer (Computer Talk, USA)

BIO

Robin Springer founded Computer Talk, a full-service assistive technology consulting firm, in 1995 to address the growing numbers of individuals with musculoskeletal injuries who required assistive technologies to remain productive. The company’s objective is to improve technological accessibility and productivity, regardless of physical ability.

Springer assesses users' needs and configures technology plans, which she implements for children and adults at all points on the spectrum of abilities. She is a columnist for Speech Technology Magazine, an advocate for individuals of varying abilities, and is currently conducting the second phase of a longitudinal study design about individuals with disabilities and their involvement in advanced degree education. She frequently contributes to publications, consults with companies on product design and implementation strategies, and lectures about assistive technologies and the disability paradigm. For more information see:

www.comptalk.com.

1. Brief description of your domain (including actual needs/requirements)

Computer Talk is a full-service assistive technology consulting firm specialising in speech recognition. Our objective is to improve technological accessibility and productivity, regardless of physical ability. We conduct feasibility studies to determine which products will best meet our clients' needs. Once the products have been identified, we implement the technology plan, configure and install the products, and provide initial and ongoing training and support. Computer Talk also advocates for individuals with "dis-abilities," educating the public as to common misperceptions about "disabilities."

2. Target group (professionals, public at large, people with disabilities, surgeons, etc.)

Our target market is any company or individual who could benefit from improved workflow design. This may be a professional who requires increased productivity or it may be an individual, with or without a "dis-ability," who wants to control his/her environment hands-free. Using our techniques, our clients can compete in the community, matching or exceeding the skills of their counterparts.

Our clients include individuals in the legal, medical, public safety, and entertainment industries, among others. Within these market sectors, some of our clients have “dis-abilities,” (or as we like to say, “varying abilities”) and require technology to increase their independence. In addition to our clients who are in the workforce, we often work with students and individuals who want to use technology primarily for home use.

In creating an environment of improved productivity, a professional may retain our services to convert to a paperless office or an office that uses speech technology instead of traditional typing and dictation services, decreasing reliance on transcription services and decreasing the dictation loop. These companies may have employees with work-related injuries who, if not for adapting the work environment, would be forced out of their jobs. The companies may have employees with varying abilities who need ancillary technology to perform their job functions successfully.

Computer Talk modifies off-the-shelf products to target the direct needs of our consumers. When necessary, we write interfaces between programmes to enable smooth interactions between them.

3. Actual modalities available in your interfaces, considering that modality = language (text, graphic, sound, 3D, etc.) + device (computer, robot, PDA, head mounted display, etc.)

We use many modalities, including speech input, speech output, haptic devices, and infrared technology in combination with traditional computers, tablet PCs, PDAs, and other hardware. We often integrate multiple products, combining them to minimise barriers, creating an accessible environment for our clients.

For example, solutions for our clients who need additional cues to compensate for lack of vision may include using a traditional computer with CCTV or large monitor, possibly also adding

speech recognition (speech input) or text-to-speech (speech output) software. For our clients who have limited mobility, we may introduce infrared or radio frequency technology with speech recognition to voice-enable one's home or office, or we may combine a head-pointing device with an on-screen keyboard and word prediction. Accessibility switches are also used in this context. Scenarios for our clients who are auditory learners may include combining text-to-speech software with word prediction or other literacy software.

The third-party software we install for our clients may be written in various languages but when Computer Talk creates our own products we predominantly use Visual Basic and VoiceXML.

4. The ideal interface in specific situations: major challenges in your domain. What interactions are envisioned in the near future (new modalities?)

One of our biggest challenges is that society tends to compartmentalise aspects of life, without acknowledging the subjectiveness of labels: a programme is "good" or "bad;" a chair is "comfortable" or "uncomfortable;" people are "disabled" or "able-bodied." These definitions are subjective, may change based on the criteria used to form the judgment, and they definitely limit us.

To combat this over-simplification, we need to change the way we look at people and situations. If we see that people are not inherently either "disabled" or "not disabled," we can understand that everyone is able-bodied and disabled, depending on the task. For example, if we are in a room and the music is so loud that it prevents us from hearing, we are as deaf at that particular time as an individual who is physiologically unable to hear. This shift of the paradigm of disability from one of inability to a spectrum of abilities creates an environment of inclusion. Once we have this sensibility it will be easier to incorporate the elements of Universal Design into products and environments.

Implementing universal design necessitates creating interfaces that can be used by as many people as possible. We do this by ensuring redundancy is incorporated in the design. Whether we are allowing a consumer to use a product by voice, keyboard, audio cues, or biometrics, we need to provide more than one method for users to access the device.

Biometrics offers many benefits to individuals who have disabilities, but there are currently security issues associated with biometrics. Using an interface in which the owner retains possession of his biometric identifiers at all times (one scenario is a contactless smart card that uses certificates to confirm the user's identity), we can provide users access to the environment without necessitating the use of one's hands or eyes to swipe a card or type a password.

5. How do you see the integration of the following aspects in current/future multimodal interfaces:

a. Context-aware adaptation

As products evolve, they should do a better job accommodating for changes in the environment; for example, a product that self-regulates when the ambient noise changes. In the context of speech recognition, this feature will benefit a wide range of users, from desktop dictation users to individuals using speech recognition in cars or at kiosks.

Broadening the availability of speaker-independent products and products that can identify a user's vocabulary/language model preference (child versus adult; oncologist versus ophthalmologist; formal writing versus casual writing) and adjust automatically, without necessitating that the user manually open and close vocabularies and profiles, will result in less cumbersome user profiles and enhanced user experience.

Devices that allow a choice of access methods, including biometrics, and that identify the user's preferred access method,

providing prompts based on those preferences, will simplify the user experience and accommodate a wider audience.

b. Continuity across media and devices and task continuity

For clients with varying abilities, it is often a challenge to master the use of an input device. If the user is required to use one interface for his home computer and a different interface for his PDA or cell phone, use of technology becomes even more daunting, and possibly impossible. We need to include redundancy in design so the user can interact with multiple devices using the same interface. The Tablet PC is an example of redundancy in design; the product includes a touch screen *and* stylus *and* keyboard capability *and* may accommodate the use of handwriting recognition and speech recognition as well as on-screen keyboard, etc.

Refining and implementing substitute access methods for hand/eye-busy situations is also becoming more widely accepted. This may include the ability to begin a task on one device and complete the task on a different device (dialog mobility) and contactless smart cards, which contain information about the user, including preferred method of access.

c. What is the best way to combine modalities in input and output?

Redundancy is key. Allowing a user to interface with the product using one's voice *or* a keyboard *or* a mouse *or* a touch screen is quite different than requiring a user to use one's voice for part of the application *and* a keyboard for another part of the application *and* a touch screen for yet another part of the application. Making a product more robust may imply, but may not ensure, that a product has redundancy in the design. If we make products robust (the product does more things) and redundant (the user can exploit the product because he/she has more ways to access the product), we make products more usable and users more efficient.

d. How to assess usability?

To determine how usable a product is, one must think about how different types of users would be able to access and utilise the product. Asking, “Could someone who is blind use this?” “What about someone who is deaf? Or in a wheelchair? Could my grandmother use this? My child?” lets us consider all aspects of the design and anticipate where users may encounter difficulties. The usability factor decreases by each question that is answered with a “no.”

e. Personal interfaces versus group interfaces (collaborative interaction)

Because personal interfaces are typically used by a single user whereas group interfaces are used by multiple often a multitude of users, when we create personal interfaces we offer the user more control over the customization of the product. This is logical because it is likely that a user of a personal interface will use the product more often and will know the product more intimately. Similarly, because a product that is used by many users may be used in shorter duration, the customisation features are not apparent or not existent. These customisation features include prosody, volume, speed, brightness, etc. In the future it may be expected that, as products begin doing a better job identifying users and their preferences, differences among systems may become more transparent.

f. Task help, task surveillance, task monitoring (for assessment):

Error logs are helpful because they provide objective information as to the cause and location of the error that has been encountered. Error logs decrease reliance on the user having to describe the problem or sequence of events that precipitated the problem, often resulting in more expedient error resolution.

With regard to programs “learning” about the user, many current products allow the user to choose whether he wants the product to learn from the session or disregard potential changes. While this may be more desirable than a system that does not learn or does not offer the choice, the feature is often used incorrectly. We may begin seeing more products that “learn smart,” i.e., the system is able to differentiate between user error (and not learn from these errors) and valid information to be included in updating the user profile, and transparently make appropriate modification. Products that incorporate this additional step will help maintain the integrity of the user profile.

g. Trust in the interface (what is the required level of confidence?)

Confidence in a product, whether it is considered “good” or “successful,” is often subjective. Factors to be considered range from cognitive or physical ability to emotional or environmental inconsistencies. As products become more intuitive they will better accommodate for these factors and will be usable by more people in more scenarios. As this happens, trust in the interface will increase.

For example, Voice over Internet Protocol, if implemented successfully, allows users to access telephonic information in a variety of mediums; users can listen to e-mail messages or read voicemail messages. Because VoIP is inherently pliant it can enable users with many types of abilities to modify the way they access telephonic information. But VoIP is not always compatible with devices such as text telephones, and incompatibility (trust in the interface) is often not discovered until a consumer attempts to use the products together. This is an example of emerging technology creating accessibility issues for people with disabilities; we have a product that’s usability may range from very usable to not usable at all. It will be interesting to see how society solves this problem; will we preclude the more than one million TTY users from using the technology or will we create standards and regulations to address the problem? This will be

the difference between defining the technology as one that limits or one that includes.

h. Interfaces for learning:

Continued improvement in the intuitiveness and ease-of-use of products will enable consumers to use products with less user training. This may be accomplished by improved natural language capabilities, improved use of nouns in speech recognition applications, and improved adaptation to external stimuli (such as sunlight interfering with infrared head pointing devices). Applications may include prompts for the user, guiding him as to acceptable responses; interactive training; and inclusion of barge-in functionality in applications that enable users to implement choices by speaking or pressing numbers on a keypad.

13.3. Academics

13.3.1. Human-Computer Interaction

13.3.1.1. Laurence Nigay (UJF, France)

BIO

PhD in computer science, Laurence Nigay is assistant professor at Université Joseph Fourier (UJF, Grenoble 1) and Institut Universitaire de France (IUF). She has been a member of the Engineering for Human-Computer Interaction group of the CLIPS-IMAG (Communication Langagière & Interaction Personne Système) laboratory since 1994.

Her research interests focus on the design and development of user interfaces. In particular her research studies centre on Multimodal and Augmented Reality (AR) user interfaces such as the component-based approach named ICARE (Interaction Complementarity, Assignment, Redundancy and Equivalence) for the development of multimodal and AR interfaces.

Laurence Nigay has received several scientific awards (including the CNRS Bronze medal in 2002 and the UJF gold medal in 2003 and again in 2005) for excellence in her research and is involved in many international scientific societies and events, as well as European research projects.

There are three main research areas that focus on multimodality:

- Human-Computer Interaction (HCI), where the computer is a tool: multimodal interaction enhances the information bandwidth between the user and the computer, and makes the interaction more robust and more flexible.

- Human-Computer Communication (HCC), where the computer is a partner: multimodal communication corresponds to an anthropomorphic approach and is based on natural multimodal human-human communication (including speech and gesture and emotion capture).

- Computer Supported Collaborative Work (CSCW), where the computer is a medium for mediated human-human communication: multimodality is conceived of as a vehicle to enhance communication between people.

Laurence Nigay quotes the lack of studies on collaborative multimodal interaction. For example, it may be interesting to compare the parallel usage of two modalities by one user with the parallel usage of one modality by two users. Nevertheless she mentions that some experiments have been carried out on collaborative multimodal interaction: the goal was to encourage collaboration amongst children for drawing tasks.

The research work of Laurence Nigay in multimodality is within HCI. It is mainly dedicated, either to the public with applications such as mobile Augmented Reality game or multimodal yellow pages on PDA, or to professionals, for example, using Augmented Reality for surgery or military aircraft.

Many modalities are treated in the context of this research:

- text for input/output,
- 3D sound for output,
- graphics on Head-Mounted Display,
- graphics on mini-screens attached to physical objects,
- speech, but for voice commands only,
- pointing devices, from the classical mouse or pen to more unusual equipment such as a pedal for surgeons or a small tactile surface fixed on the wrist (like a watch),
- embodied interaction on PDA,
- 3D gesture,
- 3D localisation tracking (of mobile users and physical objects),
- Orientation tracking (magnetometer), for example in games in order to know in which direction the player is looking.

Some of the above modalities have been developed and tested while some were simulated (Wizard of Oz techniques). In addition, some of the modalities are active (explicit commands to the system) while others are passive. Passive modalities are used to capture relevant information for enhancing execution of the task, information that is not explicitly expressed by the user to the computer such as location tracking, emotion or context capture.

For Laurence Nigay it is a major challenge to determine the accurate distribution of active and passive modalities for a given application. In any event the user must have the right to change modality and form of multimodality and be aware of the passive modalities (observability principle for privacy issue). Indeed the user must be offered the capacity to observe what is being captured about her/him (passive modalities) otherwise there

could be a breach of privacy. Nevertheless observability is also an important issue for active modalities (lexical, syntactic, and semantic feedback) so that the user is able to check if her//his multimodal inputs were correctly captured and treated.

Continuity is another important issue in multimodal interaction. It can be assessed at three levels of abstraction: Perceptual/Actional continuity, which is more closely related to physical interaction (device level); cognitive continuity, which concerns the language used for interaction; and task continuity, which deals with dialogue.

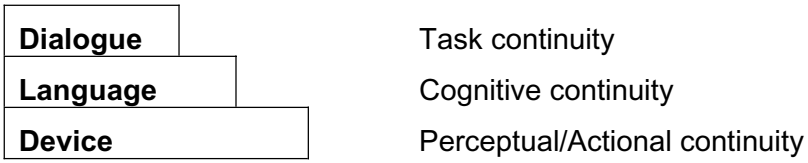


Fig. 1: Continuity

The task level has less to do with multimodality since the tasks are modality-independent as opposed to Cognitive and Physical levels, which correspond to the two levels of a modality, a modality being defined as the coupling of a physical device d with an interaction language l.

In order to assess/achieve usability of multimodal interfaces, two approaches are complementary:

- Predictive evaluation in light of ergonomic properties such as continuity as defined above.
- Experimental evaluation (multimodal corpus). Event-capture devices especially for mobile applications can be integrated in the interface in order to check how each user is interacting multimodally.

13.3.1.2. Jean Vanderdonckt (UCL, Belgium)

Jean Vanderdonckt

BIO

Master in Mathematics and Computer Science, Aggregation in Sciences, PhD in Computer Science, Jean Vanderdonckt is currently professor at School of Management (IAG), Université Catholique de Louvain (Louvain-la-Neuve), where he is the Head of the Belgian Laboratory of Computer-Human Interaction (BCHI).

Jean Vanderdonckt is the author of several papers and books; he is also co-editor-in-chief of the Kluwer International Book Series on HCI.

BCHI laboratory is involved in many research projects concerning user interface engineering, from the regional to the international level.

13.3.2. Context-aware adaptation

13.3.2.1. Joëlle Coutaz (UJF, France)

BIO : Joëlle Coutaz has been teaching at University Joseph Fourier (Grenoble, France) since 1973. She is the founder and the head of the Human-Computer Interaction Group at CLIPS (Communication Langagière & Interaction Personne Système) laboratory. She is the author of the PAC model, a conceptual software

Joëlle Coutaz architecture for interactive systems.

She is involved in many French and European research projects, as well as in international scientific societies.

HCI, the goal of which is to design and build useful and usable computers in context, is a multifaceted domain, with hardware and software aspects.

Many modalities can be involved in it: speech recognition, direct manipulations with physical tokens and surfaces, graphical and sonic rendering, etc. Their use can obviously be chosen according to the context (e.g., no speech recognition in a noisy environment).

Beyond these considerations, polymorphic embodiments of User Interfaces (UIs) can also be envisioned according to the user profile: age, skills, etc.

Interactions between a human and computer can be:

- explicit and thus consist of conscious actions performed by the user, or
- implicit, sensed and recognised by the system to support users' activities even if they are not directly intended to the system itself.

A broad range of devices is used: computers, PDAs, smart phones, but also augmented reality surfaces.

Group interfaces may exploit multimodality to support group awareness more subtly than is currently done in graphical UIs for groupware.

Augmented reality means the synergy between the physical and digital worlds: the former keeps its communication properties, but is enriched by means of digital computations.

This progress implies that the user is no longer limited to the traditional interfaces (screen, mouse, and keyboard). The UI will be able to migrate from one device to another or be distributed

among all existing resources, even though their coupling is not intrinsically provided (between Macintosh and PCs, for instance).

Many methods and frameworks to assess usability have been developed: predictive theories such as KLM (Keystroke Level Model), inspection methods of the system by human experts (using, for example, the cognitive walkthrough or usability criteria), and experimental studies either *in situ* and/or in the laboratory. Usability can be evaluated in a formative way (i.e., all the way through the development process) and/or in a summational way at the end of the project.

Augmented reality is one of tomorrow's major challenges, but it is already on the way.

13.3.3. Usability

13.3.3.1. Niels Ole Bernsen and Leila Dykbjaer (Odense University, Denmark)

Leila Dybkjaer

BIO

Respectively PhD in Computer Science and in Philosophy, Leila Dybkjaer and Niels Ole Bernsen are both Professors in the Natural Interactive Systems laboratory (NIS) of Odense, University of South Denmark.

Niels Ole Bernsen

They are involved in many Danish and European research projects regarding NIS and multimodal interfaces.

NIS (Natural Interactive Systems) and multimodal interface are somehow hard to distinguish: in order to communicate with a system in the same way they communicate together, which is the definition of NIS, people will necessarily use different modalities such as speech, body attitude and gesture.

Any way, some modalities used in multimodal interfaces are not natural, in the sense that they are not used by people in their communications (e.g. infra red).

Since it goes from medical to edutainment, both NIS and multimodal interfaces concern a wide range of users, with all kinds of needs and all levels of skills, which requires a high level of adaptability.

NIS and multimodal interface of course offer a huge potential for training, of language of course, but also for school or for learning movements, such as in sport.

Criteria for users' satisfaction and trust are hard to define. Efficiency is useless if users don't want to use the system, and a machine will never get the same trust level as a human being, no matter the degree of similarity it reaches. The question of trust, which was much discussed in the beginning, has now been reduced to basic criteria, such as the system reliability or the user feeling in control of it.

Some systems have already been implemented in this area, such as a car system that monitors the driver's preferences in order to indicate him/her the best suitable hotel according to the location, the places he/she has already been and the kind of hotels he/she likes, or a language training system helping the user to improve his/her pronunciation.

The current modalities available deal with all that concerns speech in input and output, including lip movements, with gesture, text and video.

Computers, mobile phones and PDAs can be used.

But fully NIS remains very far away; many issues still need to be dealt with.

Near future in NIS and multimodal interface will deal with handling non task-oriented conversations, multiparty conversation, speech recognition taking intonations and their semantic consequences in consideration, animated agent with gesture and face expressions.

Current issues are, among others,

- Asymmetry between the input and the output: the latter, being prepared and well thought out, is much easier to manage than the first, which involves human speech or gesture,
- Ability for the system to learn from its environment: and to react in consequence,
- Input fusion, but not only at the signal level,
- Finding accurate modalities combinations for each use of the system: modalities have to be combined in a way such that they ensure users' satisfaction. This requirement is not obvious to reach, due to the difference between theory and practice, to the various tasks to be performed by the system and to the varied categories of users. It is for example obvious that speech cannot be a modality for the seizure of a bank code.

13.3.3.2. Gitte Lindgaard (Carleton University, Canada)

Gitte Lindgaard

BIO

Professor Gitte Lindgaard is Director of the Human Oriented Technology Lab (HOTLab) at Carleton University. She holds the NSERC/Nortel/Mitel/CITO/ OCRI Chair of User-Centred Design.

She was the Principal Scientist and Head of the Human Factors Division at Telstra Research Laboratories, Australia. After a stint as Director of a private consultancy, she joined Carleton University (Ottawa) in 2000 where she is responsible for building and expanding ties between academia, industry and Government in Canada in research and HCI training.

She was Chair of CHISIG (Computer Human Interaction Special Interest Group) of the Ergonomics Society of Australia (ESA) during which she founded the OZCHI conference. She is an adjunct professor and a research associate at several Australian Universities and a Fellow of the ESA. She represents Canada on the IFIP TG13, Human Computer Interaction, and she deputy editor of the journal ”Interacting with Computers”. Her research interests include Human Computer Interaction, especially in multimedia/multimodal systems, Computer Supported Cooperative Work, and human decision-making in medicine. She has published extensively in scientific journals, from user interface design, task analysis, human judgement and decision making to pleasurable design and usability.

1. **Brief description of your domain (including actual needs/requirements)**

In our HCI we attempt to identify how interactive technology should be designed so that it meets users' needs and is easy to learn, so that it is acceptable to users, and so that it also supports more advanced users as they gain increasing experience. Our HOTLab (Human Oriented Technology Lab) is, to our knowledge, the only one that resides in a psychology department in North America, and probably also in Europe, Asia, and Australia. Since most of us are psychologists, our investigations begin and end with human preferences, biases, and performance, looking at the world from the perspective of psychologists. We want to understand, describe, and predict what it takes to first identify and then fulfil a particular human need. We also try to utilise as much as we possibly can of psychological theory, so our work is predominantly theory-driven. For example, when designing a database to support medical diagnosticians in the diagnostic process, we want to understand what kinds of difficulties diagnosticians are up against before, where the process seems to break down, and therefore understanding what information and what specific support diagnosticians need for the decision support system to offer actual support. Only then are we ready to design the user interface for the system. Most of our research lies in generating this type of understanding and then taking it further into the design of interactive technologies.

2. **Target group (professionals, public at large, people with disabilities, surgeons, etc.)**

Really, our research focuses on all of these. In the area of multisensory user interface design, we are explicitly accommodating the needs of people with different disabilities. For example, we are working with people with visual impairment in projects in which we are learning how to translate information between modalities, such as reading graphs to someone who can't see them; some work aims to identify the locus of cognitive impairment in Alzheimer's patients, particularly aiding the early

diagnosis of Alzheimer's by employing different input/output technologies.

In the context of our e-commerce-related work in which we are dissecting the notion of user satisfaction, its relationship with perceived and actual usability, trying to understanding the notion of trust, and the relationship between aesthetics, preferences, emotion, and performance, we are addressing design issues that would affect the general public. Our work in the medical arena is directed at medical diagnosticians, and our navigation research is currently framed within models of learning, especially addressing how younger people might benefit from interactive multimodal systems. Thus, we really span the whole spectrum of professionals and lay people, all age groups, and people with various disabilities.

3. **Actual modalities available in your interfaces, considering that modality = language (text, graphic, sound, 3D, etc.) + device (computer, robot, PDA, head mounted display, etc.)**

We are working with a mixture of modalities, including text, still graphics, animation, sound, and most recently, smell and tactile input and output. So far we have worked mostly with desktop equipment and also PDAs/cell phones. However, we are moving into 3-D real/virtual environment navigation as well.

4. **The ideal interface in specific situations: major challenges in your domain. What are the interactions envisaged for a near future (new modalities?)**

That is what we are trying to determine. We still don't really know in what situations one modality 'gets in the way' of, or interferes with another. For example, text + voice-over reading the text aloud simultaneously can be very annoying to the user who feels 'paced' by the voice-over whose speech may not coincide with their natural reading speed. If the voice-over says something that is related to, but is not identical with the written text, that is irritating because the user now must decide whether to read or to listen. Playing soft music to accompany the text may 'feel good'

but if the purpose of the text is for a user to learn something, especially if the message is complex, then soft music may signal that the content is easier than it is. In that case, the learner may decide to pay less attention to the textual message and run the risk of engaging in more shallow learning than the material warrants. Other examples are reading in-vehicle maps or speaking on the phone while driving.

One of our main challenges is to determine how one modality may supplement another especially in situations where the user is likely to encounter problems of information overload. For example, in network management where a user may be monitoring several types of network simultaneously, an operator might be busily fixing a problem in one network while something requiring her immediate attention is happening in another. Can we bring this to the operator's attention gently by using sound, temperature, tactile output or smell, without stressing her out at the same time? Or take another situation in which the concern is not information overload: imagine a patient with an artificial heart.

These patients must be hooked onto the power mains most of the time, but they can live on the battery for about 60 minutes before it becomes too low. Somehow we must gently make the patient aware of the status of his battery without scaring him to death. Here the problem is both one of timing and of choosing the output modality wisely.

In the short term, we are playing with using cell phones to aid visually impaired people navigate their physical environments using GPS, camera cell phones, landmarks and textual signs that the person cannot read. I believe that we will be seeing such integrations of modalities in existing technologies in the short term.

5. **How do you see the integration of the following aspects in current/future multimodal interfaces:**

a. **Context-aware adaptation**

In terms of understanding where a person is, we already have GPS; I think that will be used more and more. I have much less faith in technologies that purport to 'read' and adapt to people's moods and emotion, and frankly, I can't really see a big need for that. Systems that can sense how busy you are would be nice – doable in the short term? Not sure. I see a real need for adaptive systems that are able to recognize human information overload and poor human decisions, and that are able to override such decisions in more complex situations such as in advanced manufacturing, nuclear plants, aircraft cockpits, even in motor vehicles.

However, I also think we must be very careful in the assumptions underlying the design, implementation, and execution of such systems, especially in the consumer-oriented technologies. I am yet to meet anyone who likes the 'flexible' menus in Microsoft Word. Sure, the split menus are fine, but this growing and shrinking and swapping around of menu items according to frequency and recentness of usage are well, not to many people's liking. We also better be sure the systems work, and that they don't kick in so much that people become de-skilled and rely too much on the relevant system. That could cause havoc the moment the system breaks down.

b. Continuity across media and devices and task continuity

Yes, this is hot. People want to be able to have their email read to them when eyes are busy, like sitting bumper-to-bumper on the way home from work; they want to be able graphically to see and manipulate information presented as tables, they want to play "what if" games with all sorts of data ranging from their investments in the stock market to their company's sales data, to the frequency of occurrence of certain diseases in different age groups and geographical contexts, and they want to decide which medium to use in different circumstances I believe this is where a lot of our efforts will continue to be in the near future.

c. What is the best way to combine modalities in input and output?

Good question – that is precisely what we are spending our careers investigating. There is no “all or none” answer to that; it is highly situation-, task-, individual-, etc. dependent.

d. How to assess usability?

That is another big one. We are trying to devise methods for assessing the usability of novel mobile technologies; traditional usability methods on desktop computers carried out in a lab don’t work here. In order to understand how to assess usability of multimodal systems, I am afraid we need to be situation-specific and technology-specific until we understand the interplay between different senses, task complexity, and motivation much better.

e. Personal interfaces versus group interfaces (collaborative interaction)

I think we will see integration at both the individual and the collaborative side of interaction. We certainly need better tools for, for example annotating, even archiving and finding, images, ability to stop a training movie and ask questions, using either sound or text, and many others. I think that, particularly the translation of one visual medium (e.g. text) to another (e.g. graphs) is not too far off, and the cross-modal interaction will be more and more in demand.

f. Task help, task surveillance, task monitoring (for assessment)

Task help in the form of “just in time learning” is being addressed by a large group of researchers in the field of “Performance Technologies”, creating what they call “performance systems”. I am not on top of where this is at, but judging by the help systems I come across, it is still very rare that they actually offer any help

solving the problem(s) I experience in my day-to-day computing. I have my doubt that multimodal help will improve unless it is based on an understanding of the kinds of problems users encounter – to date I find little that encourages me on that front!

Task surveillance and task monitoring do serve a useful and legitimate purpose from the point of view of scheduling work, the number of workers, training needed, and so forth. However, they are unfortunately often being abused by management who sees monitoring as a way of policing workers' performance, even their use of email and the Internet, in ways that are unrelated to scheduling tasks and work.

Already, we have monitoring equipment with video, audio, real-time text, and database interrogation tasks being monitored. As consumers calling an organization, we are commonly told, "this call may be monitored for quality purposes". In my own work, I am yet to see these capabilities used intelligently and for the claimed purpose. Until that happens, I think that further integration aiming to make it easier for those carrying out the observations while infringing on the privacy of those being observed, is not a good idea.

g. Trust in the interface (what is the required level of confidence?)

Trust is another big issue that we are researching. Striking a balance between trusting the system and taking whatever precautions to guard against losses of files, of skill – or worse, is where I see a lot of work is needed. I have some difficulty seeing the advantage of integrating different modalities in this area, although it seems that aesthetics largely determine the way we feel about an interface and play a major role in applications for the consumer markets.

h. Interfaces for learning

Designing multimedia multimodal applications to support effective learning again boils down to understanding much more about how the different sensory modalities interact, how they supplement, or can get in each other's way. From our psychological set of biases and prejudices, it is a matter of defining the best ways to mix modalities that motivate the learner without encouraging too shallow processing of the learning material, and finding ways of portraying complex material such that the learner feels they can master it.

13.3.4. Information fusion & fission

13.3.4.1. Jean-Philippe Thiran (EPFL, Switzerland)

BIO

PhD in Electrical Engineering, Jean-Philippe Thiran is currently assistant professor and Head of the Computer Vision and Image Analysis group within ITS (Institut de Traitement des Signaux, Signal Processing Institute), one of the largest laboratories at the Ecole Polytechnique Fédérale de Lausanne (EPFL).

Jean-Philippe Thiran is involved in many national and international research projects, as well as in internationally acknowledged scientific societies and events.

1. Brief description of your domain

The Signal Processing Institute (ITS) at EPFL is one of the most important research groups in Signal Processing in

Europe and in the World, with recognised expertise in signal, image and video processing.

Our main activity is in the development of new methodologies, algorithms and technologies for signal processing, image and video representation, analysis and communication. Applications range from very low bit-rate image and video coding to multimedia content delivery, image and video analysis and to medical image analysis.

ITS is a pioneer in multimodal signal/image processing, where novel problems of signal representation and analysis appeared. We have proposed theoretical frameworks for handling multimodal signals not as several separate channels but really like a single unified signal where each channel contains marginal as well as redundant information.

Applications are found in multimodal audio-video processing, like multimodal speaker tracking or lip-reading, or in multimodal medical image segmentation and registration for image-based diagnosis or computer-assisted intervention.

2. Target groups

They can be both professionals and the public at large, when multimedia applications are involved, such as better videoconferencing or communication systems. They are also the medical community, where image-guided and computer assisted medicine is involved.

3. Actual modalities available in your interfaces

We are not primarily attached to some specific modalities. Our main research work is dedicated to conceptual research on how to extract and optimally use the marginal and redundant information contained in multimodal signal as a whole.

Therefore in that respect the actual modalities are not of importance. For some practical works we use audio and video (speaker tracking, lip reading), or multimodal medical images (image segmentation and registration). Since our mission is not to produce devices but algorithmic innovations, we cannot answer the question about devices.

4. The ideal interface in specific situations: major challenges in your domain. What are the interactions envisaged for a near future (new modalities)?

The ideal interface is able, in a changing environment, to adapt to the available modalities, and to optimally exploit the information contained in each of them.

It can be exploit the redundant information between modalities to take more robust decision and being able to move from one modality to the other smoothly, or it can be by exploiting the specific information of each modality. This really depends on the application but theories supporting this have to be developed, and this is our goal.

5. How do you see the integration of the following aspects in current/future multimodal interfaces:

a. Context-aware adaptation

This is of major importance. In a decision process, the context is a major source of information. Prior knowledge is one of the most important ways to take robust decisions. This is valid both for human beings and for computerised systems. Our work will not by such try to make the system aware of the context, but more provide the community with a theoretical framework able to best exploit the prior/context

information images, to allow the developers to build robust systems.

b. Continuity

This is also of prime importance. Multimodal signal/images contain by definition some redundant and some marginal information. The way this information is exploited has to be flexible enough to allow two opposite behaviours. Either the specific information of each modality is exploited to take more robust decisions, assuming both sources of information will always be present, or the information has to be extracted such that it is highly redundant, such that if one medium disappears, the continuity is maximally guaranteed. This is the essence of our research.

c. What is the best way to combine modalities in input and output?

If we knew the answer, we would not be working on it! The question is very valid. I believe a multimodal signal should not be considered as a set of separate streams, but as one single signal, to be processed jointly. The way it is processed depends on the application.

Although this is theoretically well justified, the practical constrains - e.g. on probability density function estimation, difference in time scales, etc. - make it difficult to really process a multimodal signal as one single signal, but requires to process the different streams separately and to fuse the decisions at the end.

d. How to assess usability?

Usability is a domain that I am not an expert of. What is sure is that usability has to be part of the whole design process, to optimally build the information-processing pipeline. Moreover, when working with multimodal interfaces, we rarely know the relative importance of each modality on the final user.

This importance should be given, in order for the designer to better weight the pieces of information brought by each modality.

13.3.4.2. Sorin Dusan (Rutgers, USA)

Bio

PhD in Electrical and Computer Engineering, Sorin Dusan is an Associate Research Professor within the Speech and Language Processing Laboratory of the Center for Advanced Information Processing (CAIP), at Rutgers University.
His main research interest is in the field of speech and language processing, and more broadly communications, artificial intelligence and biomedical engineering.

Sorin Dusan is involved in several scientific societies and events; he is the author of numerous publications.

He also participates in many research projects.

1. **Brief description of your domain (including actual needs/requirements)**

I got my PhD in electrical and computer engineering and my main interest is in research in the areas of speech and language processing. The ultimate goal of this research is to facilitate more natural, efficient, and intelligent interaction between humans and computers as well as among humans. Speech and language are at the core of human communication. Extending these capabilities to human-computer interaction will greatly enhance productivity and user satisfaction when the system performance will approach human performance in manipulating speech and language. But humans do not use the language faculty in isolation. Most of the time speech is complemented by gestures in face-to-face interactions. Multimodal interaction, centred around speech, can greatly improve the efficiency and accuracy of communication between humans and machines, while making this interaction more natural and intelligent. Depending on specific circumstances, humans fuse multiple senses and respond employing multiple channels in order to enhance the communication robustness especially in noisy or impaired conditions.

Automatic speech recognition and speech synthesis are two research directions than need to be advanced in order facilitate human-like interaction with computers and machines. However, these input and output processes need to be coordinated and fused with gesture recognition (face, hand, body) and graphical display signalling. The multimodal fusion and response need to resemble those employed by humans. My approach to automatic speech recognition and understanding, speech synthesis, and multimodal interaction is biologically inspired due to a decade of experience in biomedical engineering. I believe that these processes need to be approached in a cognitive context and not in isolation. My interest in speech (auditory) and visual perception, as well as in cognitive processes in general, is motivated by a strong belief that a better understanding of these processes can offer significant insights which could be clues in building new technologies.

2. Target group (professionals, public at large, people with disabilities, surgeons, etc.)

The impact of developing natural, multimodal interfaces centred around speech is very broad. It can affect not only the way professionals interact with computers and machines but also the way general public performs this interaction. Certain applications of speech and language processing and multimodal interaction target special categories of users such as those with various handicaps (auditory, visual, motor). About 25 years ago I developed a system that transformed the speech signal from a microphone into abstract images on a TV screen to enable the hearing impaired to “perceive” speech using a supplementary channel (visual). Later, I developed another method to enable the voice impaired after laryngectomy to produce voiced speech. Other medical applications could target the visually impaired, for example, by automatically generating synthetic speech from a web page they cannot read.

3. Actual modalities available in your interfaces, considering that modality = language (text, graphic, sound, 3D, etc.) + device (computer, robot, PDA, head mounted display, etc.)

My multimodal interface research is directed towards improving the fusion of input modalities (speech and pointing) and improving the coordinated multimodal response (graphics and speech). These speech-centred multimodal interfaces were developed for both desktop and portable computers (including PDAs). The multimodal fusion aimed at allowing users to perform flexible timing when interacting by voice and gestures. Thus, pointing is allowed before, during or after speaking.

4. The ideal interface in specific situations: major challenges in your domain. What are the interactions envisaged for a near future (new modalities?)

The human-computer interface needs to become natural, intelligent and efficient. However, these desiderates cannot be achieve without building human-like artificial agents capable of using spoken language, learning, thinking, adapting and, showing emotion. Building such artificial agents is not trivial. Artificial intelligence and cognitive science would play important roles in such an enterprise. Signal processing (speech and vision) needs to be complemented by new modules that resemble various cognitive processes that take place in the brain. On the other hand, the basic performance in automatic speech recognition, language understanding, and speech synthesis needs to be improved in order to bridge the current gap between human and computer performance. One direction is to advance the understanding of the equivalent processed that take place in the human brain.

5. How do you see the integration of the following aspects in current/future multimodal interfaces:

a. Context-aware adaptation

Context should play a much important role in human-computer interaction than it plays today. Human-human interaction is heavily context-dependent and this makes it so efficient. The adaptation of the interaction between human and machine can and should be based on context and this can be done in a few different ways.

b. Continuity across media and devices and task continuity

Cross-media continuity is possible at a few levels of representation. For an efficient interaction, it is not necessary to employ all the available modalities (input and output) with similar (redundant) information. For example, usually it is much more efficient to employ a gesture to point to an object or location than to employ speech (e.g. “select this” + pointing to the object). In other cases it is more efficient to employ speech to refer to a specific object when that object has a unique characteristics (e.g., “delete the blue cell”). Cross-devices continuity is a challenge but becomes more and more a technology issue than a research issue. Task continuity can be implemented in higher levels modules employing language understanding.

c. What is the best way to combine modalities in input and output?

There are different ways to fuse the input modalities and distribute the communication message across various output modalities. My approach is to allow the user through this fusion/distribution to interact freely and flexibly with the computer in a way similar to the way humans interact with each other.

d. How to assess usability?

Usability has not played the necessary role in the past in developing human-computer interfaces. Restricting users to interact with the machines using a specific vocabulary, modality or time slot decreases the user satisfaction and the interaction efficiency. Usability principles should dictate the design of the new interfaces and they should be rooted in the natural human-human interactions. Systems should allow flexible interactions using various modalities (sequentially or simultaneously). Usability can be evaluated by comparing real interactions with modality specific interactions.

e. Personal interfaces versus group interfaces (collaborative interaction)

Computer interfaces should allow users to personalize them (to a specific vocabulary, accent, way of timing gestures, etc.). This personalization can improve the accuracy and performance of the interaction/communication (e.g., automatic speech recognition engine can be adapted to the user's voice).

f. Task help, task surveillance, task monitoring (for assessment)

Task help and monitoring should be adaptive to the specific context. It would be useful to be available in text (graphic) form and also as a speech input modality.

g. Trust in the interface (what is the required level of confidence?)

An interface should be reliable in order to be efficient before it is natural and intelligent. Users will not consider machines to be "intelligent" until they will not be able to say if they talk to a real human agent or an artificial agent at the other end of the line.

h. Interfaces for learning.

Interfaces should have learning features whether or not used for learning.

13.3.4.3. Julien Epps (National ICT Australia)

PhD in Electrical Engineering, Julien Epps has been for two years a Researcher at the IMAGEN Programme at National ICT Australia and a Conjoint Senior Lecturer in the UNSW School of Computer Science and Engineering.

He is member of several scientific societies, involved in many events, holder of one patent and author of numerous publications.

His current work concerns multimodal user interaction, particularly speech recognition and semantic and/or statistical fusion of heterogeneous inputs.

1. Brief description of your domain (including actual needs/requirements)

Robust multimodal interfaces for practical applications, particularly interfaces combining simultaneous speech and/or video-based gesture and/or pen-based inputs. These interfaces are required for a number of different applications, ranging from control room environments to natural tabletop interaction.

2. Target group (professionals, public at large, people with disabilities, surgeons, etc.)

The main target user groups to date have been operators of control rooms (e.g. for traffic control) and users of tabletop interfaces. Control room operators perform duties under quite stressful circumstances, and often need to access information via unconventional hardware, such as large wall-mounted screens. Users of table-based hardware find the conventional mouse and keyboard interface very limiting, and hence a multimodal approach is being pursued.

3. Actual modalities available in your interfaces, considering that modality = language (text, graphic, sound, 3D, etc.) + device (computer, robot, PDA, head mounted display, etc.)

Input modalities: Speech recognition (via Bluetooth headset or far-field microphone), gesture recognition (via video-based sensors), and possibly stylus input (via e-beam on table).

Output modalities: Large wall-mounted screen or a pair of large horizontal and vertical screens in a table arrangement (rear-projected in all cases), text-to-speech synthesis, audio (beeps), and possibly avatars.

4. The ideal interface in specific situations: major challenges in your domain. What are the interactions envisaged for a near future (new modalities?)

Major challenges in my domain include interfaces that do not require the user to physically touch any input device (e.g. speech and gesture input), and eyes-busy and hands-busy interfaces. Together with colleagues, I am working on the practical and robust use of speech and video-based gesture interfaces, which I envisage will become useful for specific, narrow domain tasks in the near future. I believe that video-based gesture, still a relatively new modality, will find new applications in the near future, as new research continues to prove its feasibility for a few well-defined specific applications.

5. How do you see the integration of the following aspects in current/future multimodal interfaces:

a. Context-aware adaptation

I see context-aware adaptation as gradually pervading new interfaces, as new forms of context input become available to systems. To date, however, there have been relatively few examples of context-aware adaptation being any more essential than a ‘nice feature’.

b. Continuity across media and devices and task continuity

Discussion and awareness of this area continues to grow, however practical outcomes have been slow to follow. This is perhaps partly due to the complexity of the problems involved

and partly due to the slow migration from desktop to more mobile forms of computing.

c. What is the best way to combine modalities in input and output?

In the case of inputs, this is entirely dependent on the nature of the input modalities (i.e. whether they are correlated or redundant to some extent or whether they are complementary) and on the objective of using more than a single input modality (i.e. to permit simultaneous entry of heterogeneous input types or to enhance the robustness of a single input modality). I am less qualified to comment on combining output modalities.

d. How to assess usability?

For research prototypes, I believe that usability is still mainly about understanding users and gathering data from them in order to guess at what interface paradigms may be appropriate. For commercial multimodal systems, I believe that usability is still mainly about user acceptance of new interaction possibilities. In the future, new methods for assessing usability will be needed that account for the entirely different interfaces that a multimodal approach can allow. One method might be the extent to which a multimodal interface can adapt to any given new user.

e. Personal interfaces versus group interfaces (collaborative interaction)

Two issues spring to mind:

Evidence for the adaptation of the multimodal fusion component to the highly consistent within-user but widely varying inter-user integration patterns continues to build. For personal interfaces, I believe user experiences will improve as a result of such adaptation to individual input preferences.

A major outstanding issue for collaborative multimodal interfaces is to determine natural ways for users to indicate to the system whether any given instance of communication (including speech or a gesture) is intended for the system or for a colleague. This problem is particularly exacerbated when colleagues are remotely distributed, since the user may be simultaneously facing both the system and their colleague(s).

f. Task help, task surveillance, task monitoring (for assessment)

I am not qualified to comment on these aspects but will follow with interest the various ensuing attempts to satisfactorily resolve the difficult problem of simultaneously providing sufficient task help for novice users and allowing sufficient task freedom for advanced users.

13.3.5. Edutainment

13.3.5.1. Michael Schnaider (ZGDV, Germany)

BIO

Graduated in Computer Science, Michael Schnaider has spent five years as a researcher in the Interactive Computer Graphics Group (GRIS) of TU Darmstadt. He has now been Head of ZGDV-Department of “Visual Computing” since 2001. Created in 1984, with over 100 scientists working in its locations of Darmstadt and Rostock and its branch in Portugal, ZGDV is one of the top-ranking European research

Michael Schnaider	institutes in the sector of Computer Graphics and its applications.

It plays an intermediary role between research, training and business applications.

1. Brief description of your domain (including actual needs/requirements)

The department Visual Computing especially deals with research and development projects mainly in the area of mixed reality applications. Besides its research of technology and development of application frameworks regarding virtual and augmented reality applications, HCI for MR applications has become an essential issue. Domains considered for research range from engineering and e-work to medical and life science to education and entertainment.

Of course, the needs and requirement, which are immanent to MR applications, depend heavily on the application domain in focus. However, two generic needs can be formulated (1) the vision of creating similar sensations with Mixed reality technology than that are sensed in real environments and closely related to this (2) a similar behaviour and variety of interaction in virtual, augmented or mixed reality environments, which makes use of the full capacity of human communication.

2. Target group (professionals, public at large, people with disabilities, surgeons, etc.)

The activities of the Visual Computing department mainly target closed user groups such as professionals of particular domains. However, especially in education/training and entertainment scenarios the user groups are more open and dynamic, and basically address the large public.

3. Actual modalities available in your interfaces, considering that modality = language (text, graphic, sound, 3D, etc.) + device (computer, robot, PDA, head mounted display, etc.)

A large number of modalities are used in mixed reality applications so far. These are for the output channel: text, audio (e.g., speech output (via TTS, pre-recorded), stereo sound, surround sound), video (live streams, pre-recorded), graphics, images, 3D models, 2D animations, 3D animations. The devices used for this are single computer, computer cluster, PDA, mobile phones, various display systems ranging from monitors, (mono and stereo) projection walls, large screen LC/Plasma displays, PowerWalls, multi-sided presentation walls, CAVE to movable and/or worn displays such as HMDs or other head worn displays devices.

The input channel makes use of a number of different modalities as well. Typically, they are bound and realized on the basis of different hardware devices such as pointer (2D/3D), navigators (6DOF) or tracking systems which basically realize function like orientate, point and click or higher level functions based on these. Other more direct, human like modalities are speech input (typically spoken commands), gestures (e.g. pointing gesture), gaze tracking, and many more, which are based on Computer Vision and try to interpret human behaviour/movements.

4. The ideal interface in specific situations: major challenges in your domain. What are the interactions envisaged for a near future (new modalities?)

Within SIMILAR the domain of edutainment applications utilising mixed reality technology is mainly in focus. Here, we are typically dealing with the unspecific individual of the public and hence a large variety of expectations, how a system should behave. The ideal interface therefore should provide intuitive and most importantly natural input modalities to interact with the pieces of

information presented. In our view, kinds of interaction should be provided, which allow the user to interact without any artificial device in an unobtrusive way, as long as there is no specific need. Utilisation of dynamic gesture, body expressions, head tracking, of course conversional approaches with natural language understanding reflect best the natural way of communication. As well, interaction with physical props is also in focus.

To summarize, the ideal interface is capable of mimic the reality or in cases in which the inaccessible is presented, to provide easy-to-understand metaphors for both, presentation and behaviour that meet the user's expectations based on his/her real world experiences.

5. How do you see the integration of the following aspects in current/future multimodal interfaces:

a. Context-aware adaptation

This is an important topic! As said in 4) am important goal is to meet expectations of the users. Therefore, knowledge of the context is essential.

b. Continuity across media and devices and task continuity

In my perspective, media and devices should be seen as communication tools. Ideally, a user selects the tool most appropriate for a task in a particular situation. If the situation changed, she might like to switch to another device that provides better support in the new situation (e.g. switch from static office to mobile car/train). If we are aiming at supporting users anytime, anywhere, any-task, the continuity in question becomes important.

c. What is the best way to combine modalities in input and output?

I doubt a bit, that there is one “best way” how to combine modalities. So far approaches for early fusion or late fusion exist and to me it seems that both approaches have raison d'être. Furthermore, it is difficult to say at which level modalities should be combined, on their signal level or on their semantic level, in-between? What is really important is the effect for the user and the confidence a user can experience when communicating in a multimodal way with a technical system.

d. How to assess usability?

I believe, the well-established process of usability assessment taking into account task performance and user’s expectations of system behaviour are sufficient. However, techniques like Wizard-of-Oz should be applied first to identify the degree of acceptance. Real system trails need robust solutions without interference caused by (technical) system failures to determine requests of functional improvements.

e. Personal interfaces versus group interfaces (collaborative interaction)

There is no clear argument to favour one or the other. Personal interfaces however typically have a lower complexity compared group interfaces. For group interfaces, the understanding of the collaborative aspects of a group task/action is essential in order to identify the essential requirements to support an action.

f. Task help, task surveillance, task monitoring (for assessment)

From the technical perspective, the differentiation between a task execution and the interface to execute/control a task is important to provide solutions (the multimodal interfaces) for a large variety of tasks. Help, surveillance and monitoring however require knowledge of the task itself. Incorporating such functionality into the interfaces will limit their generalized use in different task contexts.

However, techniques available for multimodal interfaces might be used to provide the functionalities in question.

g. Trust in the interface (what is the required level of confidence?)

It is most important, that the interface and of course the underlying system behaviour are understandable and in line with real world experiences of a user. If this cannot be guaranteed, the user should be able to identify the causes why the system/interface behaves different to his/her expectations so that she/he can adapt to the required conditions (if not annoying).

h. Interfaces for learning

The domain of learning will be come more and more important. (Continuous) learning and training has already become an essential part of our lives, so appropriate interfaces are important, too. In essence, they should provide easy access to knowledge and information, and should hide any system complexity, which is not important for the learning and training task itself.

13.3.6. Medical

13.3.6.1. Richard Kitney (Imperial College, UK)

Professor Richard Ian Kitney

BIO

Richard Ian Kitney is Professor of Biomedical Systems Engineering and Dean of the Engineering Faculty, Head of the Department of Bio-Engineering (formerly department of Biological and Medical Systems) at London Imperial College. He is a fellow of several Royal Colleges, Academies and Societies in the areas of medicine, physics and engineering.

He was granted many Honours: Freeman of the City of London (1996), Fellow of the World Technology Network (1999), officer of the Order of the British Empire (2001).

The main areas Professor Kitney is currently working on are in the field of medical imaging research, which of course concerns doctors, therapists and surgeons, who will have to work with the new technologies, but also handicapped people and the public at large, who will enjoy them for better treatments. New imaging techniques will also be interesting to teach people how to move properly when they have injuries.

Professor Kitney is working on Magnetic Resonance (MR) and UltraSound (US) imaging, with computers and PDAs, among others in the scope of Computed Tomography Imaging (CT).

For what concerns MR imaging, he has been working with open magnet, General Electric new kind of interventional MR scanner,

for 6 years: this scanner was bought and installed in some various hospitals because of the radiologists' interest and the access to patients. Professor Kitney has been working in collaboration with GE and other groups such as ETH Zurich, among others for improving the Signal to Noise Ratio and developing hardware and software interacting.

The current experimentation scope, led in collaboration with biologists, concerns image guided surgery to look and repair damaged nerves in the spinal colon, through placing embryonic stem cells, living tissues, in the patient's body. These experiments have already given good results with paralysed rats, which recovered full activity within three months. This proves how nerves can build bridges by joining together to 'replace' the severed part of the spinal cord and how the brain can adapt itself to information coming from these newly built bridges.

The future is of course to develop these techniques for humans: in theory, they should work in the same way for paraplegics and quadriplegics, since part of their nerves are still fully functional.

Another research area for professor Kitney, for 18 months, is optical endoscopes, in order to have a better view of local areas: probes of 1mm diameter, with a coil at their end, for local imaging using MR. Some pretty nice results have already been obtained with this technique by doing post processing. Probes can be adapted to MR or US imaging. This very complex technology, where the antenna is inside the body, is beginning to be commercially used. They can have many therapeutic uses: treatment and reconstruction of arteries with probes which can contain up to 70 independent US elements, brain implant placement to cure Parkinson disease, liver tumour freezing through the use of a freezer probe, with sub millimetre accuracy.

Other new challenges are electro microscopy Positron Emission Tomography (PET) and optical Computerized Tomography (CT), but they are still in the gap.

All these new technologies must be conceived in the scope of operations room use, with attention on the local anatomy. They

are going to improve the training and monitoring of surgeons, by using virtual reality, post operation analysis, etc. Equipments such as haptic gloves and head up displays will be used.

13.3.6.2. Zeljko Obrenovic (University of Belgrade, Serbia Montenegro)

Dr Zeljko Obrenovic

PhD in Computer Science, Zeljko Obrenovic holds a position of researcher at the Centre for Command Information Systems and at the Laboratory for Multimedia Communications. He is also an assistant professor at the Computer Science Department of the Military Academy of Serbia and Montenegro, and a lecturer at the University of Belgrade, where he teaches several courses on HCI and software engineering. He is the author of several publications, involved in scientific societies

His current professional and scientific interests include human-computer interaction, biomedical engineering, knowledge and data modelling, software engineering, as well as development of advanced user interfaces in education and medicine. He published over 50

1. **Brief description of your domain (including actual needs/requirements)**

Our work concerning multimodal interfaces includes several areas. First of them is development of multimodal interfaces for real-time systems and for interaction with complex data sets, such as biomedical data. Multidimensional nature and complexity

of such data presents a rich field where multimodal presentation feats naturally.

Secondly, we are working on a generic framework for specification of multimodal human-computer interaction. Our modelling framework does not define any specific interaction modality but instead defines a generic approach for modelling such modalities in terms of sensual, perceptual, affective and cognitive effects they produce. Based on these ideas, we have also been working on a unified generic framework for describing user features and environmental characteristics in terms of sensual, perceptual and cognitive effects they reduce. We wanted to show that there is a fundamental connection between multimodal interface design and universal accessibility, and that awareness of these links can help both disciplines.

2. **Target group (professionals, public at large, people with disabilities, surgeons, etc.)**

Our researches have several target groups. Firstly, interfaces for presentation of complex data sets in real-time applications are primarily dedicated to the users such as physicians, surgeons, or pilots. On the other hand, our work about modelling and design of multimodal interfaces, as well as about joining universal accessibility and multimodal interaction, is primarily dedicated to HCI professionals.However, we are also trying to make our research more accessible to ordinary software engineers, by integrating HCI knowledge into their existing developing environments, such as UML tools.

3. **Actual modalities available in your interfaces, considering that modality = language (text, graphic, sound, 3D, etc.) + device (computer, robot, PDA, head mounted display, …)**

We have primarily used personal computers and PDAs, and their standard equipment. In our work with neural interfaces, we have also used special apparatus, such as EEG and EMG channels.

Beyond using standard graphical components, we employ more elaborate use of various 3D graphics and sound techniques.

4. **The ideal interface in specific situations: major challenges in your domain. What are the interactions envisaged for a near future (new modalities?)**

We are currently working on several practical projects, such as design of user interfaces for control of unmanned aircrafts. Here, we are applying ideas from our generic framework that enable treating of various user interface modalities in a unified way, taking into account their sensual, perceptual, affective and cognitive effects.

The main challenge is to create tools that can analyse the potential problems in the interaction in early stage of the development, and potentially lead toward adaptable solutions and automatic generation of user interfaces code.

5. **How do you see the integration of the following aspects in current/future multimodal interfaces:**

a. **Context-aware adaptation**

In order to create adaptive multimodal interfaces, the system has to be aware of the context in which they are used, as the semantics of modalities depends on the context. Therefore, adaptable user interfaces have to be able to use most appropriate modalities in a given context. One important topic here is providing a unified representation of modalities, user profiles, and environment characteristics.

b. **Continuity cross-media and cross-devices and task continuity**

Internet-enabled cell phones, PDAs, desktop, laptop, and wearable PCs have quite different requirements and

presentation capabilities. If we want to access this content efficiently through a broad variety of pervasive-access devices, we need to rethink how it is specified and created. These problems exist in other domains too, and therefore, it is useful to analyze pertinent methods and experiences from these domains. For example, in the software development community, there is a general need for exchanging data among tools from different vendors.

In similar way, we think that we can borrow lots from these experiences, and enable development of user interfaces that is not focused on low-level details, but on generic models of task media, and devices. The main challenge, however, is to provide automatic transformation of these high-level platform independent models, into platforms specific models used for implementation.

c. What is the best way to combine modalities in input and output?

Having in mind great diversity of input and output modalities, as well as a great diversity of usage scenarios, it is important to provide more generic, unified views on various input and output modalities, and to enable developers to analyse modalities at different levels of abstraction, with various levels of details.

d. How to assess usability?

Multimodal interfaces and researches in usability and accessibility have much in common. Treating user interfaces as multimodal systems can clearly help usability and accessibility studies, as multimodal interfaces describe human-computer interaction in terms of communication channels established between the computer and the user. Limiting environment characteristics or limited abilities of a user can be viewed as a break or decrease of throughput in these channels.

User interfaces can be described as systems that produce and exploit various effects, and constraints as limitations in usage of these effects. These unified descriptions can be used to analyse potential interaction bottlenecks, having in mind a concrete user, a user interface, and environment.

e. Personal interfaces versus group interfaces (collaborative interaction)

The introduction of "social modalities" in user interfaces, is an important issue. Despite the great amount of people who are nowadays online, the Internet lack the technical infrastructure and user interfaces necessary for communicating social cues and information. If we want to improve interaction in group interfaces, it is critical that participants perceive the social patterns of activity and relationship.

f. Task help, task surveillance, task monitoring (for assessment)

Tasks that users have to accomplish nowadays are more and more complex. Therefore, it is important to provide the users with some kind of help relative to the current task. Ideally, the system would give the user guidance with the elements that are needed for the current task, but no more than that, using interaction modalities that minimise the reaction time and errors. This is particularly important in real-time systems.

g. Trust in the interface (what is the required level of confidence?)

Trust in the interface is a complex and wide problem, and we are particularly interest in communicating trust in group interfaces and social modalities. Especially important concept here is the concept of identity, as it plays an essential role in every kind of communication. As current interfaces do not transmit many of the identity cues, it is crucial to enable interfaces to more securely communicate these identity cues.

h. Interfaces for learning

Information modes in human-computer interaction are closely related with computer-aided learning. HCI has an important role in the information process, which is mostly overlooked in learning systems. Different ways of interacting with a computer imply different ways of affecting cognitive processes closely associated with learning, and awareness of these facts could clearly improve the design of learning systems.

Glossary

Flags.

Special flags are used to express exceptions. They include: ·

[comment]: Any useful comment.·

[dangling]: A definition that has not received a common agreement yet in the consortium. ·

[obsolete]: An obsolete definition. It must be followed by a new definition.·

[ref]: The definition draws upon or replicates widely-accepted meanings. In this case, the source reference is denoted as [ref] and is listed in the "References".·

[see also]: A reference to a related concept.·

[syn]: The term is a synonym. It is followed by the term with which it is synonymous.·

Terms and definitions

Abstract Interaction Object (AIO)

(a) An abstraction of a set of CIO's with respect to a set of properties.

(b) Within the Arch software architecture reference model, it belongs to the Logical Presentation functional component. [syn] Abstract interactor, logical interactor.

Abstract interactor

[syn] Abstract Interaction Object.

Abstract User Interface (Abstract UI)

A canonical expression of the renderings and manipulation of domain concepts and functions in a way that is independent from the concrete interactors available on the targets. It is expressed in terms of workspaces (as in ARTStudio [Thevenin 01]), Presentation Units (as in SEGUIA [Vanderdonckt 93, 99] - See http://www.isys.ucl.ac.be/bchi/research/seguia.htm), or Presentations (as in TERESA [Paternò 99]).

Abstract task

In ConcurTaskTrees [Paternò 99], a task that has subtasks that belong to different task categories (e.g., subtasks that are allocated to the user and to the system).

Abstraction

(a) Process that transforms a description into a description whose semantic content and scope are richer/higher than the content and scope of the initial description content.

(b) Result of the process of abstracting. In the context of reverse engineering, elicitation of descriptions that are more abstract than the descriptions that serve as input to this process. [comment] Opposite of reification. [Amodeus 95].

Adaptability

[dangling]

(a) Ability of a UI to adapt its behaviour according to a set of predefined options.

(b) Ability of a UI to adapt its behaviour from an explicit human intervention.

Adaptable User Interface (Adaptable UI)

A UI that supports adaptability.

Adaptivity

Ability of a UI to adapt without any explicit human intervention.

Application task

In ConcurTaskTrees [Paternò 99], a task executed by the system. For example, the compilation of C code is an application task. [syn] System task.

Archetypal (domain/context of use/adaptation) model

A declarative (domain/context of use/adaptation) model that serves as input to the design of a particular interactive system.

Backward recoverability

Ability of the system to provide the user with an undo facility to return to a previous state [Amodeus 95].

Basic task

A task that can no longer be decomposed, i.e., if it were decomposed further, it would be expressed in terms of physical actions. [syn] Elementary task.

Basic window

In Trident, a logical window from which it is possible to access the windows that belong to the same Presentation Unit [Vanderdonckt 99]. [see also] Presentation Unit.

Browsability

Capacity of the system to provide the user with means to make perceivable different portions of the system functional state. (By modifying the presentation state, the user may access different portions of the state of the functional core.) [Amodeus 95].

Central domain concept

Central domain concept.

Cluster (of platforms)

A composition of elementary platforms and/or clusters. It may be homogeneous or heterogeneous

Co-domain

Range of a function. Set of possible mappings ("solutions") for a particular function. For example, the co-domain of the reification function from the abstract user interface (the domain in this case) is the set of possible concrete user interfaces. [see also] Function, Mapping.

Component

Part of a whole. It can be instantiated as a software module, subsystem, agent, interactor, abstraction, device, etc. [Amodeus 95].

Component model

Description of a component.

Composite Abstract Interaction Object

An AIO is said to be composite if it can be decomposed into smaller AIO units.

Composition of functions

The composition of the functions g: A? B and f: B? C, written as f o g, is a function such that any element of the co-domain of g corresponds to an element of the domain of f. In this definition, f and g can be composed because co-domain(g)=domain(f). The reverse composition g o f is not necessarily valid since co-domain(f) can be different from domain(g).

Compound workspace

In ArtStudio [Thevenin 01], a workspace composed of multiple workspaces. [syn] Presentation Unit (PU).

Consistency

A criteria applied frequently to increase the predictability of a UI. Consistency allows the user to generalise from specific situations to similar situations. But it is difficult to determine, at design time,

which situations a user will consider similar or dissimilar [Gram 96].

Concrete Interaction Object (CIO)

An entity of the UI that users can perceive (e.g., text, image, animation) and/or manipulate (e.g., a push button, a list box, a check box). A widget provided by a toolkit. [syn] Physical interactor, physical interaction object.

Concrete User Interface (Concrete UI)

(a) A CIO-dependent expression of the UI.

(b) In ARTStudio [Thevenin 01], a simulation of the final UI that runs only within a multi-target development environment.

Configuration

[dangling] A particular reification of presentation and dialogue models for a specific UI.

Context

An all-embracing term for which there is no consensual definition. To be operational, context can only be defined in relation to a purpose, i.e., a finality. For the purpose of the CAMELEON project, context is a short cut for "context of use". [syn] Context of Use.

Context aware UI

A User Interface that can detect changes in the context of use

Context change

A short cut for "change in the context of use" as well as for "Target change".

Context dependent (entity)

An entity, e.g., a description, whose nature is specific to a particular context of use.

Context dependent task model

A task model specific to a particular context of use.

Context independent (entity)

An entity, e.g., a description, whose nature is shared by all of the contexts of use considered for the entity

Context independent task model

A task model that integrates tasks and transitions that are common to all of the contexts of use envisioned for the system.

Context of interaction

[syn] Target.

Context of use

[syn] Target.

Context semi-dependency (of an entity)

Property of an entity, e.g., a description, whose nature accommodates more than one context of use, but not all of the contexts of use considered for that entity.

Context sensitive UI

A context aware UI that, in addition, can react to changes of the context of use. A UI that is adaptable and/or adaptive to multiple contexts of use.

Context sensitivity (of an entity)

Capacity of an entity to change the values of its attributes depending on context changes

Core configuration (of resources)

An immutable set of hardware and software resources (e.g., a laptop, a PDA, the Intel Personal Server [Want 01]).

Core resources

Software and/or hardware resources packaged as a core configuration.

Corrective decoration

A decoration used by designers to override standard options of the multi-target UI development environment. Corrective decorations are reactive.

Cross-platform UI

[syn] Multi-platform UI.

Customisability

Capacity of the user interface to support adaptability and/or adaptivity.

Decoration

Kind of information attached to a description element. A way to modify the interpretation of the description element without modifying the element per se.

Description

Any representation of a real or imagined system or entity for a particular purpose. [syn] Model.

Design recovery

[dangling] Process of recovering UI design options and models from existing source code (for example, by code static analysis, by using dynamic analysis, by behavioural analysis, by programme understanding), documentation analysis, trace examination, code instrumentation, etc. Effective design recovery implies a thorough knowledge of the domain of discourse, information external to the UI source code, and deductions from it.

Device

Physical artefact used by a system to acquire (input device) or deliver (output device) information. Examples include keyboard, loudspeaker, pen, and mouse. Shortcut for Physical device.

Device assignment

Relation between a device over a state and a non empty subset of expressions of an interaction language. A device d is assigned in state s to a set E of expressions of a language l if there does not exist any device equivalent to d over s and E. Assignment is permanent if the relation holds for any state. Assignment is total if the relation holds for E equal to the set of expressions that define l. For example, a mouse is permanently assigned to the expression of window resizing in the direct manipulation interaction language. [Amodeus 95].

Device equivalence

Relation between a non empty set of devices over a state and a non empty set of expressions in an interaction language. Devices in a set D are equivalent over a state s and a non empty set E of expressions in an interaction language L if all of the expressions of E can be elaborated using one of the devices in D. Equivalence is permanent if the relation holds for any state. Equivalence is total if the relation holds for E equal to the set of expressions that define L. For example, keyboard and microphone can be totally and permanently equivalent over natural language. [Amodeus 95].

Device redundancy

Relation between a set of devices over a state and an expression of an interaction language. Devices of a set D are used redundantly in some state s for an expression e of a language l if these devices are equivalent over s and e and they are used simultaneously to express e. For example, the user can spell a character using the microphone and type in the same character. [Amodeus 95].

Device complementarity

Relation between a set of devices over a state and a non empty subset of expressions of an interaction language. Devices of a set D are complementary over a state s and a non empty set E of expressions of a language l if E can be partitioned such that for each partition Ep of E there exists a device d of D assigned

over s and Ep. Complementarity is permanent if the relation holds for any state. Complementarity is total if the relation holds for E equal to the set of expressions of I. Language complementarity is best illustrated by spoken natural languages where concept names must be typed in. For example, in Munix, a multimodal user interface for Unix, commands that involve a file name, such as “remove”, can be expressed using the microphone for the command name and options, while file names must be elaborated with the keyboard. [Amodeus 95].

Directive decoration

Decoration used when it corresponds to rules that cannot be expressed easily in terms of general-purpose inference rules. For example, suppose the multi-target development environment includes the following generation rule: "any domain concept of type Integer must be represented as a Label in the Concrete UI". If the designer wants the temperature domain concept to be represented as a gauge, a directive decoration can be attached to that particular concept. Directive decorations are pro-active.

Dialogue Controller (DC)

In the Arch software architecture reference model [Arch 92], denotes the software component in charge of task sequencing.

Dialogue (of the User Interface)

An ordered set of physical actions between the user and the interactive system.

Dialogue model

[dangling] An abstract description of the actions, and their possible temporal relationships, that users and systems can perform at the user interface level during an interactive session [Paternò 99].

Distribution (of the User Interface)

Allocation of the user interface across the devices of a cluster of platforms. The granularity for distribution is one of (in decreasing

order): a: application level (e.g., full replication of the UI on target platforms), workspace level (e.g., windows and panels), domain concept level, and pixel level.

Distributed User Interface (distributed UI)

A user interface whose components are allocated (statically or dynamically) to different devices of a cluster of platforms.

Domain

See Mapping.

Domain concept

(a) A concept relevant to users to accomplish tasks in a particular domain

(b) An element of the domain ontology.

Domain model

A description of the domain concepts and their relationships.

Dynamic Multi-target User Interface (Dynamic MUI)

A multi-target UI that is adaptable and/or adaptive at run time. [syn] Context-sensitive UI.

Elasticity

[dangling] Capacity of a UI to support different contexts of use without requiring reconfiguration. Plasticity should be distinguished from elasticity with respect to the predefined usability properties.

Elastic UI

[dangling] UI that supports elasticity.

Elementary Abstract Interaction Object

An AIO that cannot be decomposed any further. [syn] Simple AIO.

Elementary platform

A platform from the resources of which it is not possible to compose two platforms.

Elementary task

[syn] Basic task.

Elementary workspace

In Artstudio [Thevenin 01], a workspace that cannot be decomposed further.

Enabled Task Set (ETS)

The set of tasks that are enabled over the same period of time according to the constraints indicated in the task model [Paternò 02].

Entry point

Within the multi-target reference framework [Calvary 02], any level of abstraction in the reification process from which the development of a multi-target UI is initiated.

Environment

A shortcut for Physical Environment. [syn] Physical environment.

Epilogue

(a) A closing functional portion of the execution of a reaction to context change. It includes the restoration of the execution context (e.g., resuming a suspended task)

(b) In ARTStudio [Calvary 01, Thevenin 01], a reference to a function of the Functional Core after the execution of a task.

Equivalence

A relation that is reflexive, symmetrical, and transitive. [Amodeus 95].

Evolution model

A model that specifies the actions that should be undertaken at run-time by the multi-target user interface when entering and leaving a particular context of use.

Executable User Interface

A UI ready for execution. It is expressed as compiled or interpreted code.

Extension resource

A hardware or software resource that can be added to (or removed from) a core configuration or a platform (e.g., external keyboard, mouse).

Factorization

An operation that produces from a set of target-specific descriptions of class X a new description of class X composed of a context-independent part (i.e., shared by all the targets) and context-dependent parts specific to each target.

Factorization decoration

A decoration that expresses exceptions to the nominal case used as the reference in the process of multi-targeting.

Final User Interface

The UI produced at the very last step of the reification process supported by a multi-target development environment. It is expressed as source code.

Final description

A description produced by a multi-target development environment that is not reified any further.

Final model

[syn] Final description.

Fission

(a) A computation of a process abstracting/reifying an information type into a collection of different information types to be transferred to a set of processes

(b) Decomposition of an information type at some level of abstraction into multiple information types of the same level of abstraction [Amodeus 95].

Flexibility of the User Interface

Capacity of the UI to provide users (and the system) with multiple ways of achieving tasks. Is refined into a number of properties such as multimodality, reachability, and non-preemption).

Forward engineering

The process of developing a software product. The opposite of reverse engineering.

Framework

A structure intended to guide and support some human process.

Function

If A and B are two non-empty sets, then a function f in A x B exists if it associates each element in A with one and only one element in B.

Functional Core (FC)

In the Arch software architecture reference model, implements the domain-dependent concepts and functions of an interactive system. [Arch 92]

Functional Core Adaptor

In the Arch software architecture reference model, accommodates various forms of mismatch between the Functional Core and the user interface per se of an interactive system. [Arch 92].

Fusion

(a) Computation of a process that abstracts/reifies a collection of information types received from distinct processes into a different information type to be transferred to another process

(b) Composition of multiple information types at some level of abstraction into a single information type of the same level of abstraction [Amodeus 95].

Goal

A desired modification of the current state or an inquiry to obtain information on the current state.

Grouping (of interactors)

Relationships among interactors indicating that they are logically connected.

Heterogeneous cluster

A cluster of elementary platforms whose classes are different (e.g., a cluster composed of a PC and a PDA).

Hierarchy (of interactors)

Relationship among interactors indicating that they have different levels of importance for the user.

High-level task

A task that can be decomposed into a set of subtasks.

Homogeneous cluster

A cluster of elementary platforms whose classes are identical (e.g., a cluster composed of PC's).

Honesty

A property that the presentation of the system renders its functional state appropriately (e.g., it does not distort the functional state) and in a way that is understood correctly by the user [Amodeus 95].

Hybrid multi-target User Interface (Hybrid MUI)

A multi-target UI composed of pre-computed components and dynamic components

Idempotence

Property of a function to reproduce its co-domain when applied multiple times. Some functions have the special property that applying them more than once to the same co-domain produces no further change after the first application. For instance, f(f(x)=f(x). To ensure true bi-directional engineering of a UI at any level of abstraction, the composition of all the functions involved and their corresponding inverse function should be idempotent [Bouillon 02c]. If we define rei(fi) as the reification function from the concrete interface to the final interface, abs(ci) as the reverse engineering process(abstraction function) from the final interface to the concrete UI and f as the composition of these two functions, then applying once or several times the function f to a final interface will not change the co-domain of the function f.

Initial description

Within a multi-target development environment, description provided by a human designer/developer.

Initial model

[syn] Initial description.

Interaction device

[syn] Interaction resource, physical device.

Interaction language

Language used by the user or the system to exchange information. A language defines the set of all possible well-formed expressions, i.e., the conventional assembly of symbols, that convey meaning [Amodeus 95].

Interaction language assignment

Relation between an interaction language over a state and a non empty subset of conceptual units of a system. An interaction language l is assigned in state s to a set of conceptual units C if there does not exist any interaction language equivalent to l over s and c. Assignment is permanent if the relation holds for any state. Assignment is total if the relation holds for C equals to the set of conceptual units of the system [Amodeus 95].

Interaction language complementarity

Relation between a set of interaction languages over a state and a non-empty subset of conceptual units. Interaction languages of a set L are complementary over a state s and a non empty set C of conceptual units of the system if C can be partitioned such that for each partition Cp of C there exists a language l of L assigned over s and Cp. Complementarity is permanent if the relation holds for any state. Complementarity is total if the relation holds for C equal to the set of conceptual units of the system. Language complementarity is best illustrated by co-referential expressions. For example, natural language and direct manipulation are complementary over the conceptual unit "city" and any state where the specification of a city name is possible: "flights from this city" and selection of a city name through direct manipulation [Amodeus 95].

Interaction language equivalence

Relation between a set of interaction languages over a state and a non empty subset of conceptual units of a system. Interaction languages of a set L are equivalent over a state s and a non empty set C of conceptual units of the system if all of the conceptual units in C can be represented using one of the languages in L. Equivalence is permanent if the relation holds for any state. Equivalence is total if the relation holds for C equal to the set of conceptual units of the system [Amodeus 95].

Interaction language redundancy

Relation between a set of interaction languages over a state and a conceptual unit of a system. Interaction languages of a set are

used redundantly in some state s for a conceptual unit c if these languages are equivalent over s and c, and if they are used simultaneously to represent c. For example, a wall is represented redundantly by the system via a red line (graphics interaction language) and the message "mind the red wall!" (natural language) [Amodeus 95].

Interaction object

[syn] Interactor.

Interaction resource

An input or output device used by the user to manipulate and/or observe the state of an interactive system. Examples include screens, keyboard, mouse, fingers, and real world objects (such as phicons).

Interaction space

A collection of interactors that support the execution of a set of logically/semantically connected tasks. In graphical user interfaces, an interaction space can be mapped onto a window, a set of panels. [syn] Workspace, presentation unit

Interaction task

In ConcurTaskTrees, a task performed by the user to modify and/or observe the state of the interactive system

Interaction capacity (of an interactor)

General-purpose interaction tasks (e.g., selection, deletion, navigation) that the interactor is able to support.

Interactive system

A computational system that supports a set of tasks with the participation of one or more humans.

Interactor

[dangling]

(a) An abstraction of a software component that allows users to manipulate and/or observe domain concepts and functions.

(b) A computational abstraction that allows the rendering and manipulation of entities (domain concepts and/or tasks) that require input and output resources.

Interactor Model

A description that makes explicit the properties of an interactor for a specific purpose. For example, for the purpose of multi-targeting, this description includes the representational capacity, interaction capacity, and usage cost of the interactor.

Interface model

An interface model represents all the relevant aspects of a user interface in some type of interface modelling language. Components typically included in a comprehensive interface model are user tasks, domain elements, users, presentation items, and dialogue structures. The elements of an interface model are grouped into model components [Puerta 99]. [see also] Model component.

Introspection (of an entity)

The capacity of the entity (e.g., an interactor, a software component) to export its properties and behaviour on request.

Inverse functions

Function defined by inversing domain and co-domain of a previously existing function. Forward and reverse engineering can be seen as two inverse functions since the four reification steps (used for UI production) can be recovered by their corresponding abstraction processes [Bouillon 02c]. The inverse function of a function f is denoted f -1. In this case, f is said to be inversible.

Inversible function

See inverse functions.

Level of abstraction

(a) A layer within a system whose information types are characterised by a given semantic content and scope. The lowest level of abstraction corresponds to the poorest information type with regard to scope and content.

The highest level of abstraction corresponds to the richest information type with regard to scope and content. These levels as well as any in-between level depend on the perspective or the objective of the modeller and/or the modelling technique

(b) Different perspectives in the design process of a system [Amodeus 95].

Logical interaction object

[syn] Logical interactor, abstract interaction object.

Logical interactor

[syn] Logical interaction object, abstract interaction object.

Logical Presentation Component (LPC)

In the Arch software architecture reference model, the LPC insulates the rendering of domain objects from the actual interaction toolkit of the target platform. It is expressed in terms of logical interactors. [Arch 92]

Logical window (LW)

A composite AIO or a physical window, dialog box, or panel. [syn] Elementary workspace.

Mapping

A mapping M:A?B is represented as a set of pairs (x,y) where x belongs to A, called the domain of M, written domain(M) and B is called the range of M, written range(M). A m-to-1 mapping M associates one or more in one element in domain(M) with each element in range(M).

Migration (of a User Interface)

The transfer of all, or parts, of the user interface between different devices of a cluster of platforms. May occur at run time or between sessions.

Migrability (of a User Interface)

The capacity of a UI to support migration.

Migrable (UI)

A UI capable of migration.

Nomadic application

An interactive system that supports mobile users.

Modality (in multimodal interaction)

(a) The association of a representational system with an interaction resource. For example, the associations "pseudo-natural language - microphone". "pseudo-natural language - keyboard" and "pseudo-natural language - pen" are three different input modalities for specifying commands in natural language with distinct input devices [Nigay95]

(b) One of the human perceptual senses.

Model

[dangling]

(a) Any representation of a real or imagined system or entity for a particular purpose. [syn] Description

(b) A simplified description of a complex entity or process.

Model component

Any element of an interface model. The basic components of an interface model are the task model, user model, domain model, presentation model, and dialogue model. Interface models are referred to as partial models if they include just some of the basic components and as comprehensive models if they include all of the basic components [Puerta 99].

Multi-environment targeting

The process of supporting multiple classes of environments.

Multi-environment UI

A Multi-target UI sensitive to environmental variations: It is adaptable and/or adaptive to multiple classes of environments. The users and platform classes either are modelled as archetypes or are implicitly represented in the system.

Multi-lingual UI

A UI able to accommodate variation of the natural language according to what is needed by the user. For example, the user can switch from one language to another by selecting the language from a UI menu or the system can automatically set the natural language(s) according to a preference stated in a profile.

Multimodal User Interface (Multi-modal UI)

A UI that supports multimodality.

Multimodality (of a User Interface)

Capacity of a system to support multimodal interaction, i.e., the user is provided with more than one modality (simultaneously or not) to observe the system state and/or can use more than one modality (simultaneously or not) to communicate information to the system.

Multi-platform targeting

The process of supporting multiple classes of platforms.

Multi-platform UI

A Multi-target UI sensitive to platform variations. It is adaptable and/or adaptive to multiple classes of platforms. The environment and user classes either are modelled as archetypes or are implicitly represented in the system.

Multi-target development environment

A set of tools that supports the development of multi-target UIs.

Multi-target Reference Framework

A conceptual framework that structures the development process of multi-target UIs.

Multi-target User Interface (Multi-target UI)

A user interface that supports multiple targets (i.e., multiple types of users, platforms, and environments).

Multi-targeting

The process of supporting multiple targets.

Multi-user targeting

The process of adapting to multiple archetypes of users.

Multi-user UI

A Multi-target UI sensitive to user variations. It is adaptable and/or adaptive to multiple archetypes (i.e., classes) of users. The environment and the platform either are modelled as archetypes or are implicitly represented in the system.

Observability

The capacity of the UI to make perceivable all of the domain concepts that are central/relevant to the task at hand so that the user is able to determine the state of the system [Amodeus 95].

Observed (domain/context of use/adaptation) model

An executable (domain/context of use/adaptation) model that supports the adaptation process at run-time.

Ontological model

Within the multi-target reference framework, a meta-model that makes explicit key dimensions for addressing multi-targeting. It is independent from any domain and interactive system but is intended to be conveyed in the tools used for developing and

running multi-target user interfaces. When instantiated (possibly, with tool support), ontological models give rise to archetypal models and observed models which, in turn, are specific to a particular domain and interactive system. [see also] Archetypal model, Observed model.

Ordering (of interactors)

Relationship among interactors that indicates the existence of some ordering (e.g., temporal ordering) among them.

Peripheral domain concept (for a task)

A domain concept that is not central to the task but may have an impact on it.[see also] Central domain concept.

Physical action

Action performed by either the user or the system on a physical device.

Physical device

[syn] Input device/resource, output device/resource.

Physical environment

The physical setting where the interaction takes place. It can be modelled as the set of objects, persons, and events that are peripheral to the current activity but may have an impact on the system and/or users' behaviour.

Physical interaction object

[syn] Physical interactor, concrete interaction object.

Physical Presentation Component (PPC)

In the Arch software architecture reference model, renders the domain concepts and functions in terms of physical interactors [Arch 92].

Plastic User Interface (Plastic UI)

A multi-target user interface that preserves usability across the targets, that is, the properties elicited at the design stage are

kept within a predefined range of values as adaptation to different targets occurs.

Plasticity

The capacity of a multi-target UI to preserve usability across the targets.

Plasticity domain (of a multi-target UI)

Set of contexts of use that the multi-target UI covers while preserving usability.

Plasticity threshold (of a multi-target UI)

The boundary of a plasticity domain

Platform

Set of physical and software resources that function together to form a working computational unit whose state can be observed and/or modified by a human user. It may be an elementary platform or a cluster of platforms. A shortcut for "Target Platform". [see also] Elementary platform, Cluster of platforms

Portability

Ability of a system to run on different target platforms. Covers three situations, *i.e.*, changes in the hardware resources, changes in the software resources, and moving the user to a different platform using the “same” system.

Pre-computed Multi-target User Interface (pre-computed MUI)

Results from adaptation performed during the design, implementation, or installation phases of the UI’s development : Given a functional core, a specific user interface is generated for every known target

Predicate

A boolean-valued function of the state, behaviour, or trace of a system. A predicate may represent a property [Amodeus 95].

Presentation

(a) The information provided by the user interface at a given time.

(b) The process of rendering information.

Presentation Abstract Interaction Object

An AIO whose role is to present information without allowing any user interaction.

Presentation Unit (PU)

[syn]. (a) Interaction space, workspace

(b) In TRIDENT [Vanderdonckt 93], a presentation environment required for carrying out an interactive task. It includes one or more Logical Windows and a basic window that gives access to these Logical Windows. For instance, a tabbed dialogue box is here mapped onto a PU, which is itself decomposed into LWs corresponding to the dialog box appearances depending on the active tab; conversely, a web form can be mapped onto a composite AIO in a particular LW of a given PU.

Probe (for detecting context changes)

Software mechanism that monitors and detects context changes.

Prologue

(a) Opening functional portion of the execution of a reaction to context change. It prepares the reaction: the current task is completed, suspended, or aborted; the execution context is saved; if not ready for use, the new version of the user interface is produced on the fly (e.g., a new presentation, a new dialogue sequence, etc.)

(b) In ARTStudio, reference to a function of the Functional Core before the execution of a task.

Property

An observable characteristic of a system that can be measured and described by a predicate [Amodeus 95].

Reaction (to context change)

A three-step process that permits adaptation to context changes: situation recognition, reaction computation, and reaction execution.

Reaction computation

Identification of candidate reactions to context change, then selection of the reaction that best fits the situation.

Reaction execution

A three-step process that consists of a prologue, the commutation to the new UI, and an epilogue.

Reachability

Property that allows some state or set of states to be reached from a given state through user's physical actions on the system [Amodeus 95].

Recoding (of a UI)

[dangling] Any functionally equivalent transformation of the source code of a final UI. Reformatting and refactoring are particular cases of recoding.

Reconfigurability

[dangling] Ability of a UI to support multiple targets simultaneously by offering multiple UI configurations while not necessarily preserving usability.

Recovery

The performance of actions that take a system from some “unsafe” or undesired state to one satisfying some safety property [Amodeus 95].

Recoverability

Property that the system provides the user with means to undo the effect of some action. [Amodeus 95].

Redesigning (a UI)

Changes to design characteristics. Possible changes include restructuring design architecture, altering the domain model, etc. Such changes may derive from a change of context of use.

Redocumenting (a UI)

From the UI source code, process of deriving another form of UI documentation such as, but not limited to, data structure, data flow diagram, and I/O analysis. [see also] Reformatting.

Re-engineering (a UI)

Examination and the alteration of a subject interactive system to reconstitute it in a new form and the subsequent implementation of the new form. This process encompasses a combination of subprocesses such as reverse engineering, restructuring, redocumentation, forward engineering, and retargeting [STSC]. [syn] Renovation, reclamation.

Refactoring (of a UI)

[dangling] A functionally equivalent transformation of the source code of a final UI to improve its efficiency or performance. [see also] Recoding.

Reflexivity

Reflexive functions are functions mapping an existing UI representation at a given level of abstraction to another UI representation at the same level of abstraction for the same context of use [Bouillon 02b]. In ArtStudio, reflexive functions are performed manually [Thevenin 01].

Reformatting (a UI)

A functionally equivalent transformation of a source code that changes the structure of the code to improve readability.

Regenerating (a UI)

The composition of the two reification steps from the abstract UI to the final UI. It is used to complete the process of retargeting.

To obtain another UI for any other computing platform, regenerating can be defined similarly by the composition regenf = reic o reia so as to represent the complete process by regenf o retarga [Bouillon 02b]. [see also] Retargeting.

Reification

Transformation of a description (or set of descriptions) into a description (or set of descriptions) whose level of abstraction is lower than that of the source one(s). In the multi-target reference framework [Calvary 02], the inference process that covers the inference process from high-level abstract descriptions to run-time code. Opposite of Abstraction. Relation (between interactors). [dangling] Many-to-one relationship among interactors indicating that one interactor is related to multiple interactors (for example, one interactor can enable/disable a set of interactors).

Representation multiplicity (of a domain concept)

The system's ability to offer alternative representations for a domain concept [Gram 96].

Representational capacity (of an interactor)

Types of domain concepts the interactor is able to represent (e.g., a table, an integer).

Restriction

Function obtained by restraining the domain of the initial function to those elements of the domain that satisfy a given predicate. In VAQUITA [Bouillon 02a, c, Vanderdonckt 01], constraints attached to retargeting/translation functions are defined as restriction of translations and abstractions.

See www.isys.ucl.ac.be/bchi/research/vaquita.htm/.

For a given function to be applied in a specific computing platform, there is a need to define a condition to be satisfied when applying this function. For example, a constraint may be

imposed when a translation between two computing platforms occurs, such as: "the target computing platform does not allow hierarchy of presentation elements deeper than a certain threshold". The WML language instantiates this constraint to 9 (not more than 9 presentation levels in decks and cards), while certain versions of cHTML instantiates this constraint to 4 (not more than 4 levels of cHTML tags). The restriction of function is therefore required. [syn] Selection. [see also] Retargeting.

Restructuring (a UI)

The transformation from one presentation form to another at the same level of abstraction while preserving the subject's system external behaviour (functionality and semantics) [IEEE Terminology]. The task, functions (application model), and domain models should remain identical. Moreover, the dialogue model, which is left unchanged, also represents the external behaviour.

Retargeting (a UI)

Process allowing the production of an abstract UI tailored for a particular computing platform from a final UI. Retargeting is done at design time. It is the composition of three functions, *i.e.*,: two successive abstractions followed by a translation for another platform (retarga= transa o absa o absc) [Bouillon 02c] The retargeting/translation function can be subject to restrictions due to constraints imposed by the target platform. Rather, the double abstraction up to the abstract UI level and a translation to a new context of use is independent of any computing platform.

Retasking (a UI)

The change of the task model to fit a different context of use.

Revamping (a UI)

Changing the user interface without modifying the functional core. Revamping makes it possible to modify the look and feel of a user interface considerably. Not only can the visual presentation of screens be changed, but also the phrasing can be redefined,

multimedia features can be added, and on-line documentation can be created. However, revamping does not imply a change in the requirements or re-specification. Revamping is a useful re-engineering strategy when an organisation wishes to adopt graphical user interfaces (GUIs) by using middleware products, which sit between the legacy system and the UI. The new UI is created manually [Csaba 97] or automatically [Stroulia 00]) at design time. Revamping is frequently performed by designers to beautify a presentation according to users' needs.

Reverse engineering

(a) Analysis of a software system so that the software is more understandable for maintenance, evolution, and re-engineering purposes

(b) Analysis of a system to identify its components and their dependencies to extract and create system abstractions and design information. The original system is not altered. However, additional knowledge about the system is produced. Opposite of "forward engineering".

Robustness (of the User Interface)

The capacity of the UI to prevent user and system errors, as well as the UI's ability to increase the chance of successful task accomplishment. [Amodeus 95].

Selection

[syn] Restriction. [comment] A selection is the same as a restriction (mathematical term), but is more frequently used in database engineering.

Situation recognition

Identification of the current context of use.

State

Assignment of values to names representing the observables of a system [Amodeus 95].

State vector (of a component, of a system)

Names that define the state of a component or system [Amodeus 95].

System Task

In ConcurTaskTrees, [syn] Application task.

Target

Triple of the form "e, p, u" where e is an element of the environments set considered for the interactive system, p is an element of the platforms set considered for the interactive system, and u is an element of the users set for the interactive system.

Target change

A change of at least one element in the triple "e, p, u".

Target environment

The archetypal set of environments envisioned for the interactive system.

Target platform

The archetypal set of platforms envisioned for the interactive system..

Target aware (UI)

[syn] context aware UI.

Target sensitive (UI)

[syn] context sensitive UI.

Target user

The archetypal set of end-users envisioned for the interactive system.

Task

(a) A goal, together with some procedure or set of actions that will achieve the goal.

(b) An activity that should be performed in order to reach a goal.

Task category

In CTT, definition of how the performance of the task is allocated (task category can be user, system, interactive, or abstract).

Task domain concept

A concept identified by task analysis as relevant to the user to accomplish tasks in that domain. [syn] Domain concept.

Task model

Description of a set of tasks and their relationships.

Task operator

An operator that denotes relationships between tasks.

Task presentation set

Set of tasks supported by one presentation.

Task type

Indication of the semantic effect obtained by the performance of a task (e.g. selection, show information, etc.).

Transient description

[syn] Transient model.

Transient model

Within a multi-target development environment, an intermediate description used in the process of producing multi-target UI's.

Transition task

When performed, a basic task that triggers a new presentation.

Translation

The operation that transforms a description intended for a particular target into a description of the same class but aimed at a different target.

User Interface

The software component of an interactive system that allows users to observe and manipulate domain concepts. In the Arch model, it is composed of the Functional Core Adaptor, Dialogue Controller, Logical Presentation Component and Physical Presentation Component [Arch 92].

UI dialogue

See Dialogue of the UI.

UI distribution

See Distribution of the UI.

UI migration

See Migration of the UI.

Task migration

Dynamic transfer of task performance between agents (whether these agents are humans or computational).

Transition UI

Feedback provided to the user during the adaptation of the UI to changes of context of use.

Usability

The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [ISO 91].

Usage cost (of an interactor)

Measures the system and human resources that the interactor requires.

User

A shortcut for "target user".

User model

A description of a set of elements that characterises user preferences and knowledge. It can be dynamically updated.

User task

In ConcurTaskTrees, a task performed by the user without the interactive system (i.e., an internal cognitive activity, such as making a decision).

Workspace

[syn] Interaction space.

References

[Amodeus 95] Salber, D., Coutaz, J., Nigay, L., Faconti, G., Paterno, F., Duke, D., & Harrison, M., "The System Modelling Glossary", Amodeus project document SM/WP26, 1995. http://www.mrc-cbu.cam.ac.uk/amodeus

[Arch 92] Arch, "A Metamodel for the Runtime Architecture of An Interactive System", The UIMS Developers Workshop, SIGCHI Bulletin, 24(1), ACM, 1992.

[Bouillon 02a] Bouillon, L., Vanderdonckt, J., & Souchon, N., "Recovering Alternatives Presentation Models of a Web Page with VAQUITA", Chapter 27, Proceedings of 4th Int. Conf. on Computer-Aided Design of User Interfaces CADUI'2002 (Valenciennes, 15-17 May 2002), Kluwer Academics Pub., Dordrecht, 2002, pp. 311-322.

http://www.isys.ucl.ac.be/bchi/publications/2002/Bouillon-CADUI2002.pdf

[Bouillon 02b] Bouillon, L., Vanderdonckt, J., & Eisenstein, J., "Model-Based Approaches to Reengineering Web Pages", in Proceedings of 1st International Workshop on Task Models and Diagrams for user interface design

TAMODIA'2002 (Bucharest, 18-19 July 2002), Academy of Economic Studies of Bucharest, INFOREC Printing House, Bucharest, 2002, pp. 86-95.

[Bouillon 02c] Bouillon, L. & Vanderdonckt, J., "Retargeting Web Pages to other Computing Platforms with Vaquita", Proceedings of the IEEE 9th Working Conference on Reverse Engineering WCRE'2002 (Richmond, 28 October-1 November 2002), IEEE Computer Society Press, Los Alamitos, 2002, to appear. http://www.isys.ucl.ac.be/bchi/publications/2002/Bouillon-WCRE2002.pdf

[Calvary 01] Calvary, G., Coutaz, J., & Thevenin, D., "A Unifying Reference Framework for the Development of Plastic User Interfaces", in Proceedings of IFIP WG2.7 (13.2) Working Conference EHCI'2001 (Toronto, May 2001), M. Reed Little & L. Nigay (Eds.), Springer Verlag Publ., LNCS 2254, pp.173-192.

[Calvary 02] Calvary,G., Coutaz, J., Thevenin, D., Limbourg, Q., Souchon, N., Bouillon, L., Florins, M., Vanderdonckt, J., "Plasticity of User Interfaces: A Revised Reference Framework", in Proceedings of 1st International Workshop on Task Models and Diagrams for user interface design TAMODIA'2002 (Bucharest, 18-19 July 2002), Academy of Economic Studies of Bucharest, INFOREC Printing House, Bucharest, 2002, pp. 127-134.

[Csaba 97] Csaba, L., "Experience with User Interface Reengineering – Transferring DOS panels to Windows", in Proceedings of 1st Euromicro Working Conference on Software Maintenance and Reengineering CSMR'97 (Berlin, 17-19 March 1997), IEEE Computer Society Press, Los Alamitos, 1997.

[D1.1 02] Calvary, G., Coutaz, J., Thevenin, D., Bouillon, L., Florins, M., Limbourg, Q., Souchon, N., Vanderdonckt, J., Marucci, L., Paternò, F., & Santoro, C., "The CAMELEON Reference Framework", Deliverable D1.1, CAMELEON Project, 2002.

[Gram 96] Gram, C. & Cockton, G. (Eds.), "Design Principles for Interactive Software", IFIP WG 2.7 (13.4), Chapman & Hall Publ., London, 1996.

[IEEE Terminology] IEEE Terminology in Reverse Engineering, IEEE.
www.cc.gatech.edu/reverse/bibliography/terminology.html

[ISO 91] ISO9241-11 Ergonomic requirement for office works with VDT's – guidance on usability. Technical report, International Standard Organisation, 1991. [Nigay 95] Nigay, L. & Coutaz, J., "A Generic Platform for Addressing the Multimodal Challenge", in Proceedings of ACM Conf. on Human Aspects in Computing Systems CHI'95 (Denver, May 1995), ACM Press, New York, pp. 98-105.

[Paternò 99] Paternò, F., "Model-based Design and Evaluation of Interactive Applications", Springer Verlag, Berlin, 1999.

[Paternò 02] Paternò, F. & Santoro C., "One Model, Many Interfaces", Chapter 13, Proceedings of 4th Int. Conf. on Computer-Aided Design of User Interfaces CADUI'2002 (Valenciennes, 15-17 May 2002), Kluwer Academics Pub., Dordrecht, 2002, pp. 143-154.

[Puerta 99] Puerta, A.R. & Eisenstein, J., "Towards a General Computational Framework for Model-Based Interface Development Systems", in Proc. of ACM Conf. on Int. User Interfaces IUI'99 (Los Angeles, January 1999), ACM Press, New York, 1999, pp. 171-178.

www.arpuerta.com/pubs/iui99.htm

[Stroulia 00] Stroulia, E., Thomson, J., & Situ, Q., "Constructing XML-speaking wrappers for WEB Applications: towards an Interoperating WEB", in Proceedings of the IEEE 7th Working Conference on Reverse Engineering WCRE'2000 (Brisbane, 23-25 November 2000), IEEE Computer Society, Los Alamitos, 2000.

[STSC] www.stsc.hill.af.mil/reng/defin.asp

[Thevenin 01] Thevenin, D., "Adaptation en Interaction Homme-Machine: le cas de la Plasticité", PhD. thesis, Grenoble, France, 2001. http://iihm.imag.fr/publs/2001

[Vanderdonckt 93] Vanderdonckt, J. & Bodart, F., Encapsulating Knowledge for Intelligent Automatic Interaction Objects Selection, in Proc. of the ACM Conf. on Human Factors in Computing Systems INTERCHI'93 (Amsterdam, 24-29

April 1993), S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel & T. White (Eds.), ACM Press, New York, 1993, pp. 424-429.

www.acm.org/pubs/articles/proceedings/chi/169059/p424-vanderdonckt/p424-vanderdonckt.pdf

[Vanderdonckt 99] Vanderdonckt, J., & Berquin, P., "Towards a Very Large Model-based Approach for User Interface Development", in Proc. of 1st Int. Workshop on User Interfaces to Data Intensive Systems UIDIS'99 (Edimburg, 5-6 September 1999), N.W. Paton & T. Griffiths (éds.), IEEE Computer Society Press, Los Alamitos, 1999, pp. 76-85.

www.isys.ucl.ac.be/bchi/publications/1999/Vanderdonckt-UIDIS99b.pdf

[Vanderdonckt 01] Vanderdonckt, J., Bouillon, L., & Souchon, N., "Flexible Reverse Engineering of Web Pages with VAQUITA", in Proceedings of IEEE 8th Working Conference on Reverse Engineering WCRE'2001 (Stuttgart, 2-5 October 2001), IEEE Computer Society Press, Los Alamitos, 2001, pp. 241-248.

www.isys.ucl.ac.be/bchi/publications/2001/Vanderdonckt-WCRE2001.pdf

[Want 01] Want, R. & Shilit, B., "Expanding the Horizon of Location-Aware Computing", IEEE Computer, Vol.