Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

INS

Information Systems

INformation Systems

INS

Vocabularies for description of accessibility issues in multimodal user interfaces

Ž. Obrenović, R. Troncy, L. Hardman

REPORT INS-E0702 JANUARY 2007

Centrum voor Wiskunde en Informatica (CWI) is the national research institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organisation for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3681

Vocabularies for description of accessibility issues in multimodal user interfaces

ABSTRACT

In previous work, we proposed a unified approach for describing multimodal human-computer interaction and interaction constraints in terms of sensual, motor, perceptual and cognitive functions of users. In this paper, we extend this work by providing formalised vocabularies that express human functionalities and anatomical structures required by specific modalities. The central theme of our approach is to connect these modality representations with descriptions of user, device and environmental constraints that influence the interaction. These descriptions can then be used in a reasoning framework that will exploit formal connections among interaction modalities and constraints. The focus of this paper is on specifying a comprehensive vocabulary of necessary concepts. Within the context of an interaction framework, we describe a number of examples that use this formalised knowledge.

1998 ACM Computing Classification System: K.4.2 {Computers and Society}{Assistive technologies for persons with disabilities} H.5.2 {Information Interfaces and Presentation}{User Interfaces} D.2 {Software}{Software Engineering}
Keywords and Phrases: Multimodal interaction, universal accessibility, inclusive design, formal models

Vocabularies for Description of Accessibility Issues in Multimodal User Interfaces

Željko Obrenović, Raphaël Troncy, Lynda Hardman
CWI, P.O. Box 94079, 1090 GB, Amsterdam, The Netherlands
`Firstname.Lastname@cwi.nl`

Abstract

In previous work, we proposed a unified approach for describing multimodal human-computer interaction and interaction constraints in terms of sensual, motor, perceptual and cognitive functions of users. In this paper, we extend this work by providing formalised vocabularies that express human functionalities and anatomical structures required by specific modalities. The central theme of our approach is to connect these modality representations with descriptions of user, device and environmental constraints that influence the interaction. These descriptions can then be used in a reasoning framework that will exploit formal connections among interaction modalities and constraints. The focus of this paper is on specifying a comprehensive vocabulary of necessary concepts. Within the context of an interaction framework, we describe a number of examples that use this formalised knowledge.

Keywords: Multimodal interaction, universal accessibility, inclusive design, formal models

1 Introduction

The long-term goal of our research is to use formal models of multimodal user interfaces and interaction constraints to allow the (semi-)automatic analysis of required human functionalities and anatomical structures for a particular (multimodal) interface. Figure 1 illustrates the basic theoretical framework for our approach: we describe multimodal user interfaces as systems that communicate a message, an effect, by means of a modality stimulating a particular human functionality or anatomical structures, such as, sensory, motor, perceptual or cognitive. On the other hand, constraints describe influence on various factors on human anatomical structures and functionalities. For example, a simple text presentation engages many visual perceptual functions, such as shape recognition, visual grouping by proximity, grouping by good continuation, as well as other cognitive and linguistic functions. Interaction constraints, such as user disability or environmental conditions, reduce or completely eliminate some of the effects. For example, users with a central field loss disability cannot read text at usual font sizes in usual lighting conditions. By combining these descriptions, it is possible to see if the designed interface will be appropriate for a specific situation, and it enables adaptation of user interfaces according to user profiles and situational parameters. With our approach, developers can concentrate on more generic effects, providing solutions for different levels of availability of specific functionalities or anatomical structures. In this way, it is possible to create adaptable solutions that adjust to user features, preferences and environmental characteristics, Obrenovic and Starcevic (2004); Obrenovic et al. (2007).

From a developer's point of view, an advantage of this framework is that it is possible to design more flexible and more reusable solutions, aimed at a broader set of situations. Most previous work on designing solutions for people with disabilities focuses on a specific set of disabilities, or on specific situations, Abascal (2002). Bearing in mind the diversity of disabilities and situations, it is clear that development and maintenance of such systems is complex and non-optimal. An advantage of a unified description of user features, preferences and environmental characteristics, is the potential for reusing solutions, created for a particular disability, for non-disabled users in situations that limit the interaction in the same way. As well as providing more universal

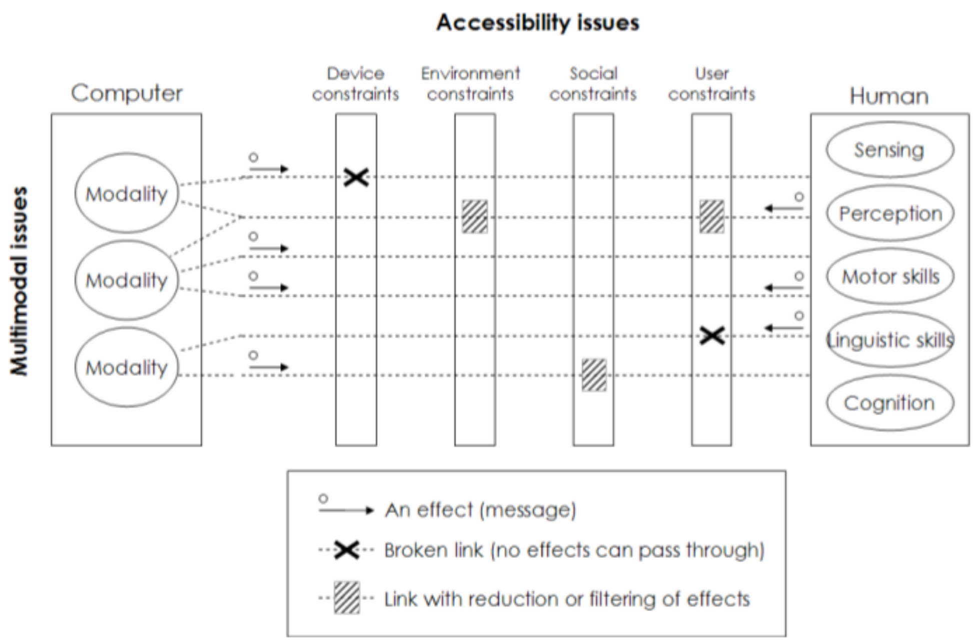

Figure 1: Modalities, constraints, and effects from Obrenovic et al. (2007). Computers and humans establish communication channels over which they exchange messages (effects) that engage a subset of human functionalities and anatomical structures. Modalities produce these effects, while various interaction constraints reduce or completely eliminate some of these effects.

solutions, this could also solve a number of ethical problems, since the design concerns effects and their constraints, rather than the term 'disability', which often introduces negative reactions. Indeed, constraints are often not a consequence of user physical limitations. For example, when interacting with a computer while driving a car, the driver is in a similar situation as a user with limited vision. These situations thus do not have to be treated differently, and solutions from one domain can be reused in another domain.

In the following, we present the basic idea of our approach, and discuss the main topic of the paper: the definition of a comprehensive set of vocabularies as a formal description of modalities and constraints (section 2). We present then some simple use cases where we have used terms from our framework to describe a concrete user interface and the human functionalities and anatomical structures required (section 3). Finally, we conclude the paper and outline some future work (section 4).

2 Vocabularies for Describing Accessibility Issues

A central problem for describing accessibility issues in multimodal interfaces is the definition of a vocabulary for the description of interaction effects in terms of human functionalities. Such a vocabulary would provide terms for describing abstract models of multimodal interaction. In this respect, our approach is similar to existing work in the area of abstract user interface representations, such as User Interface Markup Language (UIML), Extensible Interface Markup Language (XIML), W3C XForms and Alternate Interface Access Protocol (AIAP), Trewin et al. (2003). These abstract models define a vocabulary of modeling primitives for describing elements of user interfaces. Several research groups have tried to improve Web accessibility by adding annotations to web pages to help users understand the meaning of the information as opposed to its presentation and order, Bechhofer et al. (2006). Researchers have also emphasised the importance of user information and its relationships to device profiles, Velasco et al. (2004), discussing vocabularies

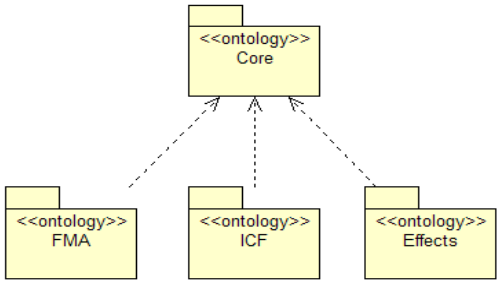

Figure 2: Ontologies for Describing Accessibility Issues

that should be used for description of these profiles. However, many of these solutions mostly focus on abstracting existing user interface platforms and content description, i.e. they are closer to implementation technology, and they do not provide a vocabulary for describing important accessibility issues and human factors involved in interaction. Our goal is to add semantics about accessibility issues and human factors. A similar attempt to defining vocabularies used for description of multimodal interaction has been taken by Ole Bernsen (1994). In his Modality theory, he introduces a generative approach to the analysis of modality types and their combinations, based on his taxonomy of generic unimodal modalities of representation. In this theory, each interaction modality is described in terms of five properties: linguistic (yes/no), analog (yes/no), arbitrary (yes/no), dynamic (yes/no) and media (visual/audio/haptic). In our work, however, we advocate a more generic solution, which enables describing human factors involved in multimodal interaction with more details and with standard vocabularies.

Instead of creating a new vocabulary from scratch we exploit two existing resources (Figure 2): *The International Classification of Functioning, Disability and Health* (ICF)[1] and *The Foundational Model of Anatomy* (FMA)[2]. While these resources cover the description of human functionalities and human anatomy comprehensively, they lack descriptions of interaction effects at the required level of granularity. To compensate for this, we also propose our own vocabulary for describing interaction effects in a multimodal environment.

These separate vocabularies need to be conceptually integrated as well as expressed in a language that can be processed within a system. The main contribution of this paper is the formalisation and conceptual integration of these three resources. We use the Web Ontology Language (OWL)[3] as the main language for describing these vocabularies, which allows us to use the large number of existing suites of knowledge management technologies and tools. In the following, we first describe each of the three vocabularies and then discuss how they can be combined.

2.1 The International Classification of Functioning, Disability and Health (ICF)

The International Classification of Functioning, Disability and Health (ICF) defined by the World Health Organisation provides a comprehensive overview of many important functions of humans. The ICF is a good candidate for describing all important human functionalities, as it provides a detailed description of human functions structured around the following broad components:

- body functions and structure,
- activities (related to tasks and actions by an individual) and participation (involvement in a life situation), and
- information on severity and environmental factors.

[1] `http://www3.who.int/icf/onlinebrowser/icf.cfm`
[2] `http://sig.biostr.washington.edu/projects/fm/`
[3] `http://www.w3.org/2004/OWL/`

In ICF, functioning and disability are viewed as a complex interaction between the health condition of the individual and the contextual factors of the environment as well as personal factors. The picture produced by this combination of factors and dimensions is of "the person in his or her world". The classification treats these dimensions as interactive and dynamic rather than linear or static. ICF has, however, several shortcomings that complicate its formalisation, including[4]: contrasting classifications, confusion between classes of activities and their qualities or features, incorrect and incomplete classifications, and over-simplification or over-emphasis of parts. Nevertheless, this resource is widely used in the health community, and by providing some connection with it, we are able to use the same standard terminology, and possibly reuse medical profiles described in these terms.

We have formalised part of the ICF ontology as an OWL ontology. Currently we have included only the concepts required by the ICF checklist[5] since they provide a good summary of the content of the whole classification. The formalization of 160 concepts reproduces the *is-a* hierarchy given in the checklist, distributed on four levels where the top-level concepts are `ActivitiesAndParticipations`, `BodyFunctions`, `BodyStructures` and `Environment`.

2.2 The Foundational Model of Anatomy (FMA)

Another useful resource is the Foundational Model of Anatomy (FMA). FMA represents a coherent body of explicit declarative knowledge about human anatomy. It is developed and maintained by the Structural Informatics Group at the University of Washington. It has also been formalised in OWL by the medical informatics group at Stanford. However, although we can directly use FMA, it misses many important concepts about human functionalities, as they cannot be described by anatomical properties. OWL release of FMA is available at: `http://webrum.uni-mannheim.de/math/lski/release.html`.

2.3 Interaction Effects

ICF and FMA provide a number of concepts about human functionalities and anatomy, but they still lack terms for more detailed description of effects that some modalities produce. For example, with ICF, we can specify that an interaction modality requires human visual perception, and FMA can provide us with a description of all parts of the human perceptual system, but none of these resources provides terms for describing details, such as, if it is expected that users perceive grouping, highlighting, or three-dimensional position of the objects. Furthermore, there are also different ways how perceptual grouping, highlighting or three-dimensional perception can be achieved. To overcome this problem, we have created a simple taxonomy of interaction effects not covered by ICF or FMA, Obrenovic and Starcevic (2004). This vocabulary describes additional sensory, motor, perceptual, and cognitive effects, from resources such as Gestalt psychology. We have formalised this resource as an OWL ontology that contains 114 concepts, some of which are shown in Table 1.

2.4 Combining the Vocabularies

In order to use the various vocabularies together, we need to connect them. Wache et al. (2001) reports three ways for doing so, namely the *single ontology*, the *multiple ontologies* and the *hybrid* approaches. In the first approach, all the vocabularies are merged in a single global ontology, while in the second one, an additional representation formalism defining the inter-ontology mapping is needed. The hybrid approach, which we have adopted, considers both aspects: separate vocabularies co-exist and are linked using a core-level ontology.

Figure 3 presents the basic concepts defined in our core ontology, with the relations to the key concepts from the three vocabularies described above. This ontology extends our previous

[4] See: `http://ontology.buffalo.edu/medo/ICF.pdf`
[5] `http://www3.who.int/icf/checklist/icf-checklist.pdf`

Grouping	3D cue
Gestalt visual grouping	Visual 3D cues
Grouping by similarity	Stereo vision
Grouping by motion	Motion parallax
Grouping by texture	Linear perspective (converting lines)
Grouping by symmetry	Relative size
Grouping by proximity	Shadow
Grouping by parallelism	Familiar size
Grouping by closure	Interposition
Grouping by good continuation	Relative height
Highlighting	Horizon
Gestalt visual highlighting	Audio 3D cues
Highlighting by color	Inter-aural time (or phase) difference
Highlighting by polarity	Inter-aural intensity (or level) difference
Highlighting by brightness	Head Related Transfer Functions (HRTFs)
Highlighting by orientation	Head movement
Highlighting by size	Echo
Highlighting by motion	Attenuation of high frequencies
Highlighting by flicker	
Highlighting by depth	
Highlighting by shape	
Audio highlighting	
Highlighting by intensity	
Highlighting by pitch	
Highlighting by rate	

Table 1: Some perceptual effects defined in the interaction effects ontology

proposal of interaction modalities, Obrenovic and Starcevic (2004), and interaction constraints, Obrenovic et al. (2007). The integration of the three vocabularies described above with our core ontology is finally available at: `http://www.cwi.nl/~media/ontologies/multimodality.owl`.

We introduce the concept of human entity, which describes an anatomical structure, or a function. An interaction modality can then be described in terms of the human entity it requires for interaction. An interaction constraint is defined in terms of the human entities that it restricts. The FMA ontology provides a number of concepts for describing human anatomical structures. The ICF body structure concepts provide a similar, but less detailed, classification of human anatomical structures. The FMA and ICF body structure concepts overlap, but FMA provides much more comprehensive and better formalised data. For the description of human functional artifacts, ICF provides concepts for the description of body functions, and functions related to human activity and participation. Neither of these, however, allows more detailed description of many parts. Our interaction effects ontology fills this gap by defining additional functional entities at sensory, perceptual and cognitive levels. These three vocabularies together provide sufficient coverage of human functionalities and effects to allow the types of mappings we envisage between the available functionalities and appropriate modalities.

In addition to this coverage of description, we also need mappings between the different vocabularies. Currently, the only relation among concepts from different vocabularies is the human entity concept. For a more elaborate analysis, more relations among concepts are necessary, for example, by establishing a mapping between the ICF body structure and the FMA concepts. These relations are also necessary to enhance the description. For example, if we describe that a user is not able to process a sound, it means that not only the sensory, but also all the audio perceptual effects will not be appropriate for that user. If the user cannot use the central visual field, limitation of vision processing will affect all visual perceptual effects, as well as linguistic effect of reading. In a similar way, low colour processing will decrease the use of the highlighted colour

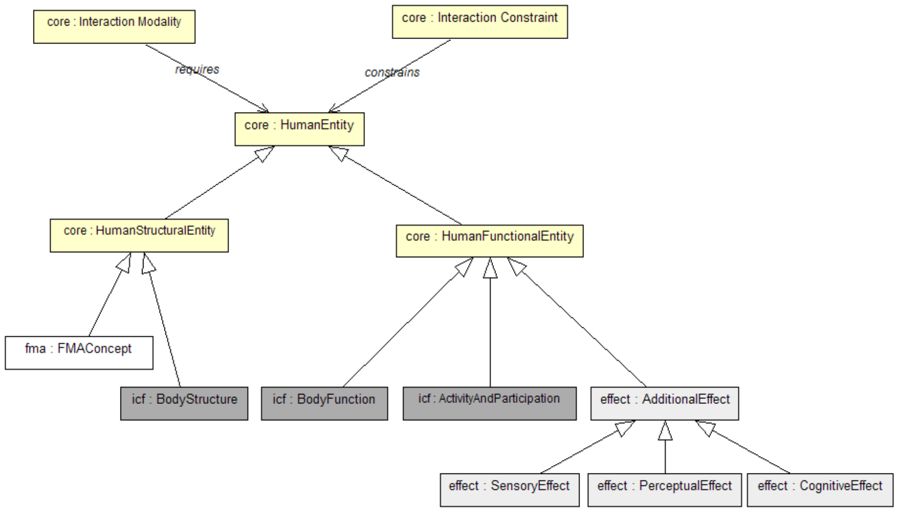

Figure 3: The core ontology, with relations to the key concepts from the three vocabularies

effect, while contrast processing will reduce shape recognition and highlighting by brightness.

3 Describing Accessibility Issues with the Vocabularies

In this section we present a number of examples that illustrate how we can use the vocabularies to describe accessibility issues in multimodal user interfaces. First we show how the vocabularies can be used to describe the requirements of standard interaction modalities. Then we show an example of description of (implicit) design decisions. In the end, we present the descriptions of interaction constraints.

3.1 Describing Interaction Modalities

We describe interaction modalities in terms of the human entities they require in order to enable interaction. These descriptions can provide richer semantics about many implicit requirements of interaction modalities. For example, figure 4 shows a simplified description of human entities required by the aimed-hand movement, a modality often used in graphical user interfaces. Aimed-hand movement is a complex modality that integrates hand movement input with visual feedback. To describe these modalities, we need concepts from all three vocabularies. Hand movement input modality, such as those used to control the mouse, requires human hand anatomy (described with concepts from the FMA ontology), and no impairments in mobility of joints, muscle power and muscle tone, as well as absence of involuntary movements (described with concepts from the ICF ontology). Visual feedback requires user eye and visual context (concepts from FMA ontology) and seeing and attention functionalities (concepts from ICF ontology). In addition we describe additional perceptual functions introduced by visual feedback: highlighting by motion and shape of the cursor, and optionally with the depth if the cursor has shadow (these concepts are defined in the interaction effects ontology).

Figure 5 shows a simplified description of interaction requirements of speech interaction. This is a complex modality that integrates speech input and output (defined relative to the computer). On an anatomical level (described with FMA concepts) speech interaction requires human vocal tract (stomatognathic system) for user speech, ear, auditory cortex and auditory additional cortex. On a functional level (described with ICF concepts) speech interaction requires functions of speaking, voice, hearing, receiving spoken message, language, conversational, general voice and speech functions, and usage of short term memory.

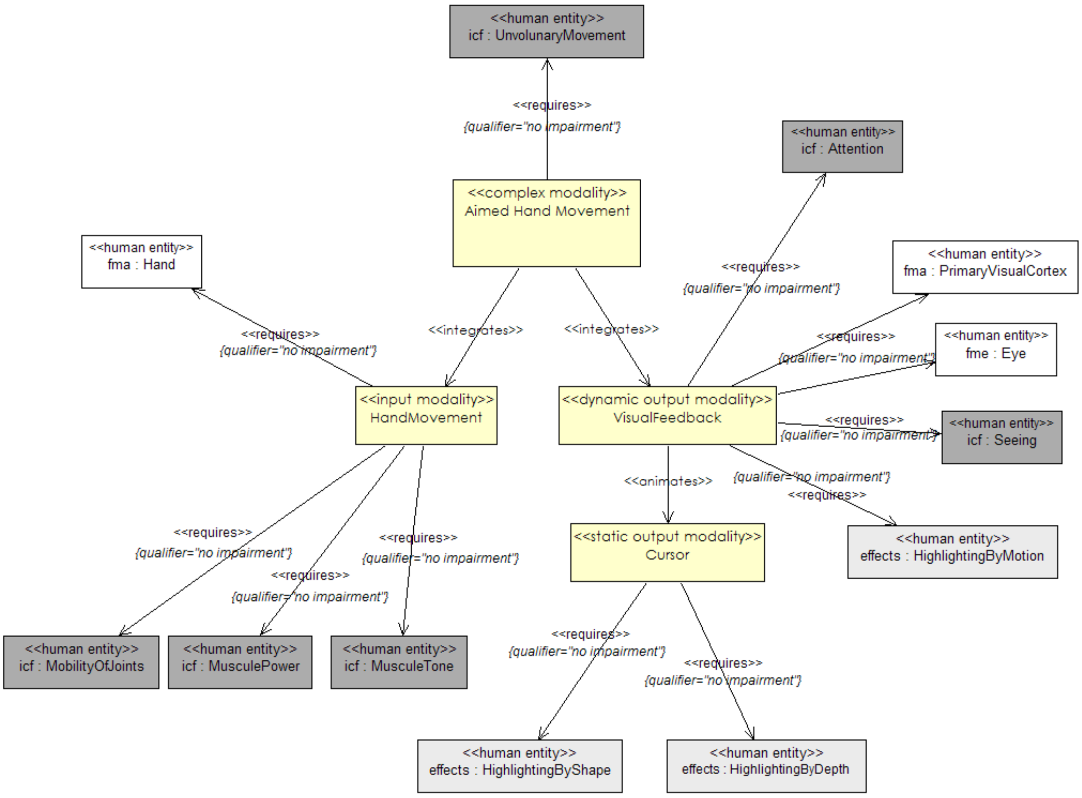

Figure 4: Description of human entity required by aimed hand movement

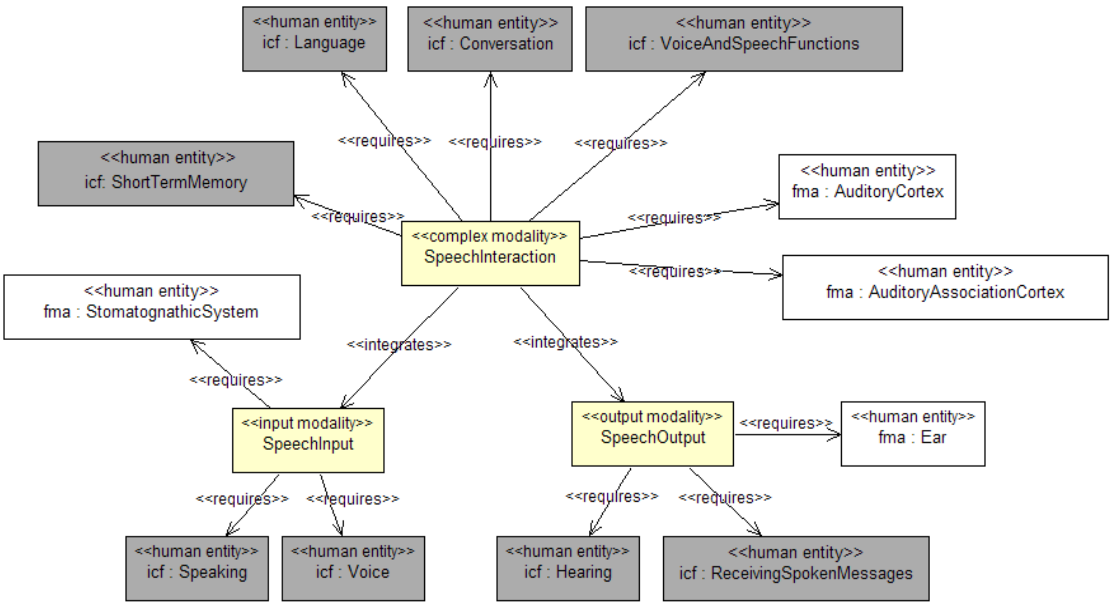

Figure 5: Description of human entities required speech-interaction

Figure 6: Simplified images of a part of MultimediaN e-culture interface

More complex modalities, such as those that use three-dimensional presentation, can also be described in this way, Obrenovic and Starcevic (2004). We can combine these descriptions with descriptions of interaction constraints, such as those described in section 3.3, to see if particular modalities can be used in a given context. We can also use descriptions of interaction modalities to identify potential conflicts in requirements. For example, Karl et al. (1993) found that the use of speech to issue commands interfered with short-term memory requirements that constituted part of the experimental task. Applying our proposed modeling of combinations of modalities allows us to select those with non-conflicting requirements.

3.2 Describing Designers' (Implicit) Decisions

User interfaces can be viewed as one-shot, higher-order messages sent from designers to users, Prates et al. (2000). In designing a user interface, the designer defines an interactive language that determines which messages will be included in the interaction. However, multimodal user interfaces are usually implemented with commercially available implementation platforms, which do not integrate the concepts of modality and multimodal integration. As a result, it can be impossible to determine the designer's original intent, which can be important when analysing and reusing parts of the user interface. The vocabularies that we have presented can be used to describe some of these intentions, enabling a designer or an HCI expert to state their aims and accessibility requirement of the interface.

Figure 7 shows an example of how we can describe design intentions of a particular interface. The figure focuses on description of perceptual effects used in a part of the MultimediaN user interface shown in Figure 6. The interface shows an ordered list of images, with titles and names of authors. Even though we describe a simple part of the interface, there are lots of important implicit elements of this presentation. Images, image titles, and author names are perceptual entities grouped by proximity, in order to be perceived as a whole. Image title and author names are linguistic modalities, requiring user reading and knowledge of language in order to be understood. Image title is also a hyperlink, visually highlighted by a colour, and by flicker when a mouse cursor is moved over it. Image presentations are grouped in a line in order to exploit perceptual effect of grouping by good continuation, and by similarity of their shape. Images are sorted so that the user exploits left-to-right perception of ordering.

This example shows many high-level effects used in the interface. It also illustrates the need for concepts not present in ICF and FMA. With ICF and FMA, in this case, we can only say things that are common for graphical user interfaces, i.e that the interface requires human eye, visual cortex, user visual perception, and function of reading.

3.3 Describing Interaction Constraints

Constraints are associated with a set of human entities that they restrict. As we describe interaction modalities and constraints using the same vocabularies, we can combine these description with description to identify potential interaction problems or select modalities that are not affected by the constraints.

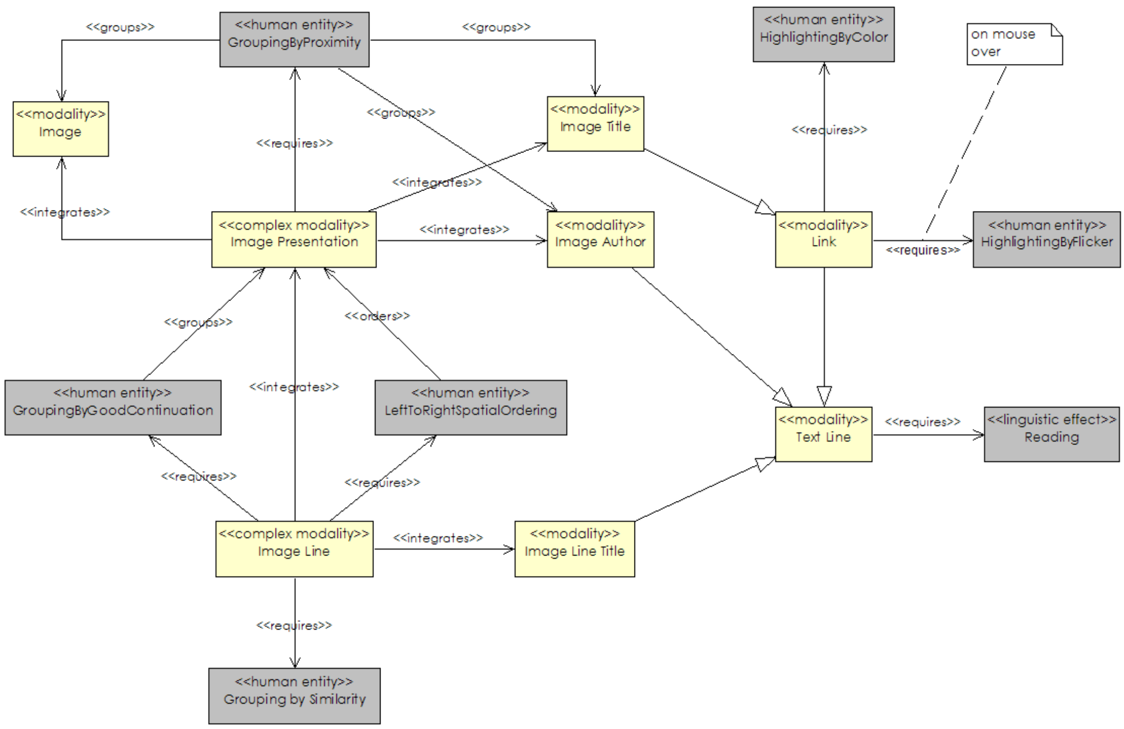

Figure 7: UML diagram describing modalities and effects used in a MultimediaN interface

Figures 8 and 9 show UML models of two interaction constraints: central field loss disability, and noise.

Central field lost is a disability that limits fovea processing to a very low level (Figure 8). In terms of FMA, this means that this modality constraint fovea centralis element. In terms of human functions, the disability constraints ICF functions of seeing, watching, reading, writing, learning to read and write, and in general many perceptual functions.

Noise in the environment primary constraints human speech and audio interaction (Figure 9). In terms of FMA, this means that this modality constraint effects associated with the ear. In terms of human functions, affect ICF functions of speaking, hearing, receiving spoken messages, conversation, and attention, and interaction effects of audio highlighting and grouping.

4 Conclusion and Future Work

There are many steps that have to be taken to achieve our long-term goal of using formal models of interaction modalities and interaction constraints to build solutions that can automatically analyse accessibility issues. First step is the definition of comprehensive vocabulary for formal description of modalities and constraints. When such vocabulary is present, even in a simple form, it is possible to improve the design of multimodal user interfaces in many directions. The main benefit of models created with such vocabulary is an explicit representation of accessibility issues using standard terms. Explicit representation lead to more automation, while using standards for knowledge representation, we automatically inherit the possibility to reuse many existing knowledge analysis tools.

Our next step is the definition of a reasoning framework that can exploit the semantics from descriptions of interaction modalities and constraints (Figure 10). The basic idea of the framework is that applications define the context by providing descriptions of user interfaces in terms of interaction constraints, and description of user, device and environment profiles in terms of interaction

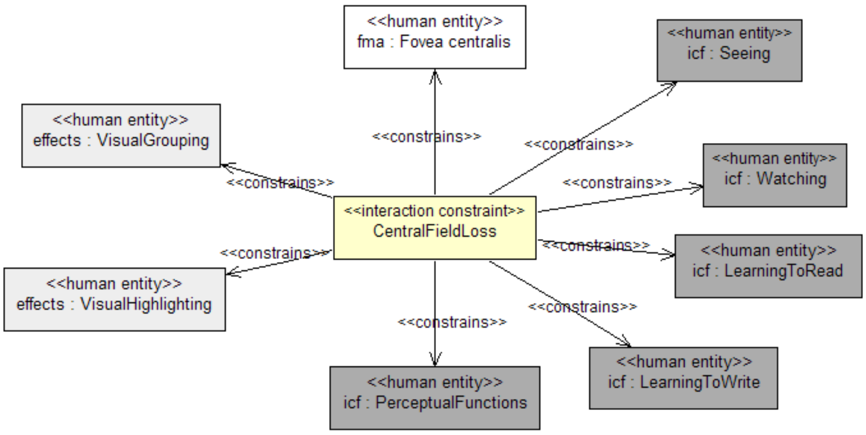

Figure 8: Description of central field loss disability (fovea vision loss)

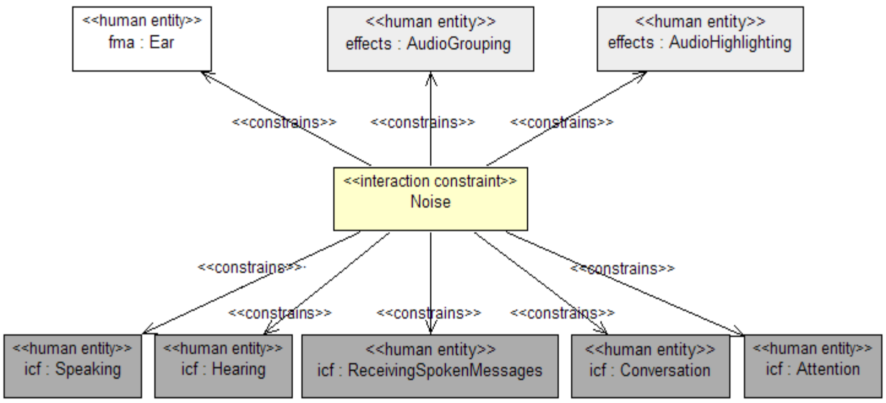

Figure 9: Description of influence of noise

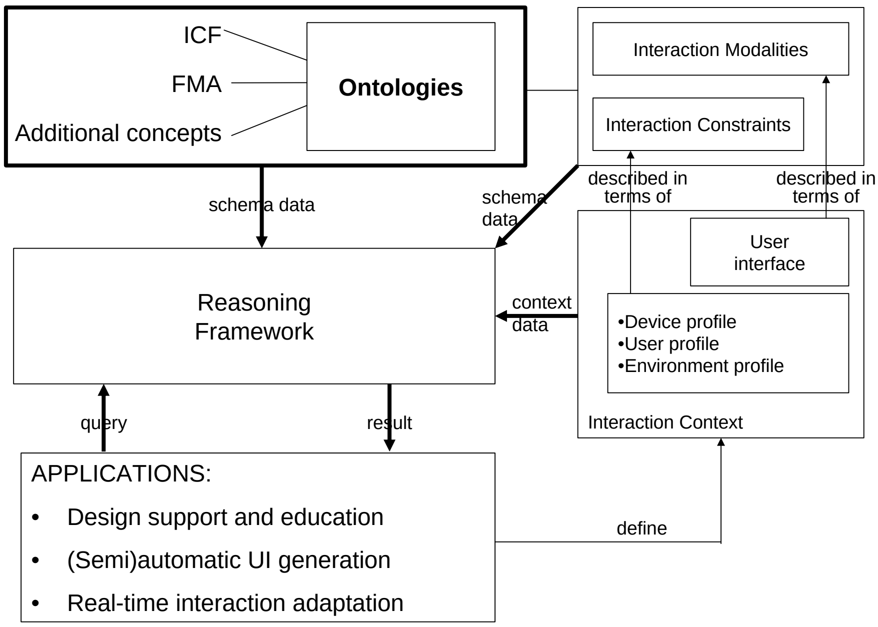

Figure 10: Using the descriptions of accessibility issues

constraints, and then use the framework to reason over these data and semantics relations. The reasoning framework can be used as a design support and education tool, enabling designers to verify their high-level decisions, and explore relations among concepts. Systems that generate user interface can use it to select appropriate modalities, or change them in real-time. The proposed framework can also be a good basis for approaches, such as user interface adaptation and content repurposing, that tackle the problem of developing content for various users and devices. The main idea of our approach is that existing content can be analysed in order to create higher-level description using the concepts from our ontologies. If original content is not appropriate for the user or situation, we can try to repurpose it into a new form, changing improper modalities, but trying to keep higher-level effects contained in the user interface.

Acknowledgments

The research leading to this paper was partially supported by the European ITEA Passepartout Project and by the European Commission under contract FP6-027026, "Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space". We would like to thank Alia Amin, Joost Geurts and Jacco van Ossenbruggen who provided useful feedback on the work described here and whose suggestions significantly improved this article.

References

Abascal, J. (2002). Human-computer interaction in assistive technology: from "Patchwork" to "Universal Design". In *IEEE International Conference on Systems, Man and Cybernetics*.

Bechhofer, S., Harper, S., and Lunn, D. (2006). SADIe: Semantic Annotation for Accessibility. In *5th International Semantic Web Conference (ISWC'06)*, pages 101–115, Athens, GA, USA.

Bernsen, N. O. (1994). Foundations of Multimodal Representations: A Taxonomy of Representational Modalities. *Interacting with Computers*, 6(4):347–371.

Karl, L., Pettey, M., and Shneiderman, B. (1993). SpeechActivated versus Mouse-Activated Commands for Word Processing Applications: An Empirical Evaluation. *International Journal of Man-Machine Studies*, 39(4):667–687.

Obrenovic, Z., Abascal, J., and Starcevic, D. (2007). Universal accessibilty as a multimodal design issue. *Communications of ACM (to appear)*.

Obrenovic, Z. and Starcevic, D. (2004). Modeling Multimodal Human-Computer Interaction. *Computer*, 37(9):65–72.

Prates, R. O., de Souza, C. S., and Barbosa, S. D. J. (2000). Methods and tools: a method for evaluating the communicability of user interfaces. *Interactions*, 7(1):31–38.

Trewin, S., Zimmermann, G., and Vanderheiden, G. (2003). Abstract User Interface Representations: How Well do they Support Universal Access? In *2nd ACM Conference on Universal Usability (CUU'03)*, pages 77–84, Vancouver, BC, Canada.

Velasco, A., Mohamad, Y., Gilman, S., Viorres, N., Vlachogiannis, E., Arnellos, A., and Darzentas, S. (2004). Universal access to information services - the need for user information and its relationship to device profiles. *Universal Access in the Information Society*, 3(1):88–95.

Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., and Hübner, S. (2001). Ontology-Based Integration of Information - A Survey of Existing Approaches. In *Workshop on Ontologies and Information Sharing*, pages 108–117.