Human-Computer Interaction -- INTERACT'03
M. Rauterberg et al. (Eds.)
Published by IOS Press, (c) IFIP, 2003, pp.535-542

Using Ontologies in Design of Multimodal User Interfaces

Željko Obrenović, Dušan Starčević, Vladan Devedžić
University of Belgrade, FON - School of Business Administration,
Department of IS, POB 52, Jove Ilića 154, 11000 Belgrade, Yugoslavia
Email: {obren, starcev, devedzic}@fon.bg.ac.yu

Abstract: In this paper we introduce an ontology-based approach to the design of multimodal user interfaces. The main benefit of our approach is the unified ontology of multimodal interaction, where we have integrated the knowledge and common concepts from different domains of multimodal interaction in a uniform way. The unified ontology of multimodal interaction can be viewed as a tool that a designer of user interfaces can use to see relations between his design and various human factors. Using this ontology we have explored novel design approaches in order to provide means for abstract definition of a human-computer interaction at different levels. We have also integrated the ontology with existing user interface platforms, what allows mapping of high-level concepts from the ontology to primitives in a user interface platform, and evaluation existing user interfaces in order to see what perceptual or cognitive effects they produce.

Keywords: multimodal user interfaces, human factors, design, ontologies.

1 Introduction

Although the processor power of our computers has increased exponentially, user interfaces have not been able to make use of this trend. The channel between a human and a computer has not changed significantly for a long period, and at this time it is a bottleneck for many applications. In order to make interaction between humans and computers more efficient and natural, multimodal user interfaces have to engage human perceptual, cognitive, and communication skills (Reeves, 2000). Even though there have been many research activities in this area, we still lack solutions and approaches that can better exploit natural human capabilities. One of the reasons for this is a cross-disciplinary nature of HCI. A development of multimodal user interfaces requires knowledge from many fields such as computer sciences, cognitive sciences, and biological sciences, but many of the HCI engineers and designers do not have good background knowledge outside computer sciences. A designer of multimodal interfaces has a difficult task of finding relevant information about human factors of interest for his/her work searching through many different sources. Another problem associated with cross-disciplinary nature of HCI are possible misinterpretations among researchers from different fields. Various communities of experts and practitioners examine problems from different angles and are concerned with different dimensions of the problem. These people may use different terminology, but they can mean pretty much the same.

This problem has been recognized earlier (Philips, 2001), but not until recent years have knowledge management technologies allowed for easy and efficient sharing of our knowledge. Current state-of-the-art knowledge management technologies such as ontologies are suitable for achieving this task (Devedzic, 2002). An ontology is usually defined as an explicit specification of an abstract, simplified view of a world we want to represent. It specifies both the concepts inherent in this view as well as their interrelationships. In our case, an ontology can give us a common and standardized language for sharing and reusing knowledge about phenomena in the multimodal interaction field.

In this paper we introduce an ontology-based approach to the design of multimodal user interfaces. The central role in our approach has the unified ontology of multimodal interaction, where we have integrated the knowledge and common concepts form different domains of multimodal interaction in a uniform view. We have used this ontology as a foundation for exploration of novel design approaches for multimodal interaction systems that could provide us with means for definition of multimodal interfaces at different levels of abstraction. In the next section we give a short overview of sources of knowledge used in our ontology. After that we describe the basic idea of our ontology-centered approach. The section that follows describes in more details the ontologies we have developed. Then we give examples of applying our ontology to various user interface design problems.

2 Sources of Knowledge

HCI is an extremely multidisciplinary field that requires knowledge from many sources. We have identified three main sources of this knowledge:

- Medicine and biology,
- Cognitive sciences, and
- Computer sciences.

Medical and biological sciences can give us useful knowledge about human anatomy and about human sensory and motor physiology such as resolution and sensitivity of different human sensors (Boff, 1986). We can also use the knowledge about human physiological signals detectable outside a human body, such as electromyograph (EMG), electroencephalograph (EEG) (Jovanov, 1999), and electrooculograph (EOG). This knowledge is valuable in many new applications such as electrophysiological interactive computer systems or brain-computer interfaces (Allanson, 2002), and can be very important for disabled users.

Cognitive sciences can give us information about more complex human mechanisms such us perception, memory or attention. For example, processing of depth cues and motion happens at the pre-attentive perceptual level, which means that there is no cognitive load for the user. Careful exploitation of these capabilities with 3D graphics and animation can make multimodal presentation much more effective.

In last two decades computer science researchers have proposed and implemented various solutions for creating integrated multimodal interfaces in many application domains. The solutions include various taxonomies, toolkits and performance studies (Blattner, 1996).

The virtual reality (VR) community has accomplished relevant research and performance studies that can be very useful in multimodal user interfaces. VR involves many of human-factors issues including human performance efficiency in virtual worlds, task characteristics, user characteristics, human sensory and motor physiology, new design metaphors, health and safety issues and the social impact of the technology (Stanney, 1998). But in order to bring these results into the game we need more formal and standardized representation of the results, since currently they are mostly available in a textual form.

3 Ontology-based approach

We have used *the unified ontology of multimodal interaction* as a starting point for exploration of novel design approaches for multimodal interaction systems. Consequently, in our research we have focused on two main topics:

- The design of the unified ontology of multi–modal interaction, where we have integrated the knowledge and common concepts form different domains of multimodal interaction in a uniform view;
- Applying of the unified ontology of multimodal interaction to the design of multimodal inter–faces.

The unified ontology of multimodal interaction can be viewed as a tool that a designer of user interfaces can use to see relations between his design and various human factors. This ontology provides a common and standardized language for sharing and reusing knowledge about phenomena from various domains relevant for the design of multimodal user interfaces. Our aim was to identify a basic set of concepts from each of the domains and to establish relations among them. Speaking practically we wanted to achieve the following (Devedzic, 2002):

- Precisely defined terms and better structured definitions of multimodal interaction domain concepts, not just text-based information;
- High expressiveness, enabling us to efficiently describe each aspect of multimodal interaction;
- Coherence and interoperability of resulting knowledge bases, using standard modeling and storage technologies;
- Scalability of ontologies providing us with means for abstract definition of interactions at different levels.

We have used our unified ontology as the basis for exploration of a novel design approaches for multimodal interaction systems in order to provide means for abstract definition of a human-computer interaction at different levels. The unified ontology of multimodal interaction gives primitives for creating relatively high-level models of multimodal interfaces where we can use concepts such as mode, integration of modes, and perceptual effects.

Most of current interfaces are described at a relatively low-level of abstraction using primitives such as text boxes, labels or canvases. In order to make our unified ontology more practically feasible we have mapped the concepts from the unified ontology to the primitives available in existing user interface platforms. Connecting our ontology with these platforms makes it possible to map high-level concepts from our ontology to one or more primitives in a user interface platform, or to evaluate existing user interfaces in order to see what perceptual or cognitive effects they produce.

3.1 Expected benefits

The uniform view of our ontology provides valuable knowledge about important relations among HCI concepts. Unification of different views on the field can give us a “fresh view” that could raise new ideas and approaches. The unified ontology of multimodal interaction can provide the context of multimodal concepts where we could perceive many relations that are not always obvious.

In addition, the ontology provides us with terms that could be used for standardizing descriptions of multimodal interfaces and results of experiments. In this way the ontology could be used as a basis for collaborative further development of knowledge about phenomena in HCI. Accordingly, the first step in using of this ontology could be the reintegration of existing solutions and experimental results in a new broader framework.

Mapping of concepts from existing platforms to the concepts from the unified ontology can provide us with means for creation of tools that can automate some phases of the design and evaluation of new multimodal interfaces.

Finally, it is important to note that by using ontologies we inherit the possibility to use other knowledge management technologies such as data mining, which can help us in finding interesting patterns in data. Also, the proposed solution represents a good basis for new user interface design approaches such as an agent-based approach.

4 Ontologies

We have used several tools to develop our ontology. We have specified all ontologies in RDF Schema (RDFS) and RDF format using the Protégé 2000 ontology editor (Friedman-Noy, 2001). RDFS is used for description of classes (concepts), while RDF is used for description of instances of these classes. For example, we have described many input and output devices in RDF files using terms defined in input and output device ontologies encoded in RDFS files.

We have also made UML models of all ontologies. We have used UML as the main notation for the description of our models. The UML is an open standard and it has standard mechanisms for defining extensions for specific applications contexts such as ontology modeling, and we have used these mechanisms extensively in our approach.

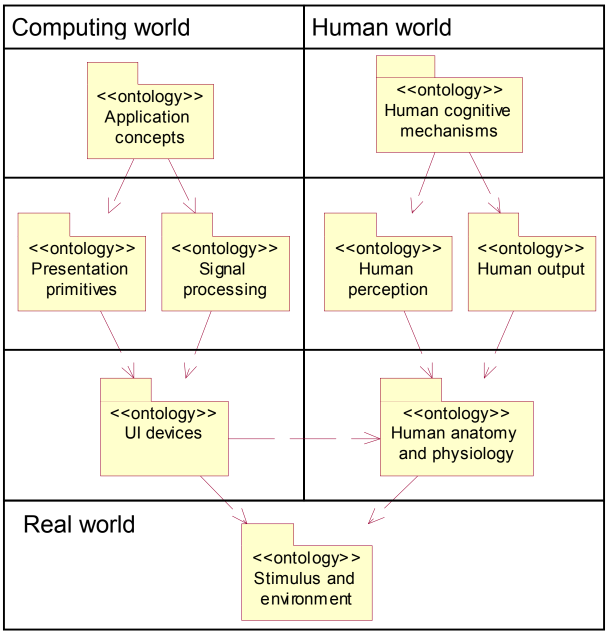

Figure 1. Relations among the domain ontologies of multimodal interaction represented using the UML package notation.

Our ontology is a complex ontology composed of many interconnected ontologies organized in four levels as shown in Figure 1. Each of the ontologies describes one domain of interest. This modular organization of the ontology is more flexible than the all-in-one approach since the knowledge is grouped into smaller sections that are easier to maintain and extend. In the following sections each of this ontologies will be described in more details.

4.1 Ontologies of human factors

There are four ontologies of human factors: the human anatomy and physiology ontology, the human perception ontology, the human output ontology, and the human cognitive mechanisms ontology.

In the ontology of human anatomy and physiology we have described the anatomy of human body and human sensual system. Since these topics involve a significant amount of literature in biology and medicine, our basic goal was to create sufficient but much simpler view on a human body. This ontology can give valuable information about various physical properties of human body. For example, this ontology can be used to describe the bandwidth of the human muscle groups and limbs used for manipulation of an input device (Card, 1991). Sensory part of the ontology can be used to describe elements of human sensory apparatus involved in processing a stimulus. In addition, our ontology defines connections of sensory regions and muscles with appropriate parts of cortex regions. This data could be useful in detail EEG recording of some brain regions, or for elimination of EEG artifacts that result from muscle activity. This is especially important for new electrophysiological and brain-computer interfaces.

Human output represents a relatively high-level description of human actions and effects that a computer can detect. Human output is classified into four main categories: *human visual output*, *human audio output*, *human mechanical output*, and *human physiological output*. We have divided human visual output into three groups: *gestures*, *tooled output* and *biometric output*. Human audio output is classified in two groups: speech and non-speech output. Human mechanical output describes various human motor actions such as movement, twisting, grasping or releasing. Human mechanical output can be simple, or it can be a series of actions.

In our ontology of human perception we have identified perceptual concepts that are often used in contemporary user interfaces. These perceptual mechanisms include *pattern recognition*, *grouping*, *pop out*, and *perception of three-dimensional cues*. These mechanisms are further specialized in visual, audio, and haptic perception mechanisms.

The ontology of cognitive mechanisms includes several of the mechanisms that were recently used in HCI research such as spatial memory, attention, and curiosity. For example, spatial memory can be effectively used in order to create better learning environments for users, as psychology research tells us that people remember spatially distributed information based on its location relative to their body, as well as the environment in which the information was learned (Tan, 2001).

4.2 Computing ontologies

There are four computing ontologies: the application domain ontology, the signal processing ontology, the presentation primitives ontology, and the user interface devices ontology. The application domain ontology defines concepts such as task, task characteristics, as well as forms of interaction between the application and the user: direct manipulation, agent based or mixed initiative. The signal processing ontology defines concepts that describe various forms of signal integration including interfaces for signal processing components such as speech recognition, machine vision, and neural networks. The presentation primitives platforms includes various higher-level models of standards UI platforms such as graphical user interface (GUI), 2D primitives and 3D primitives.

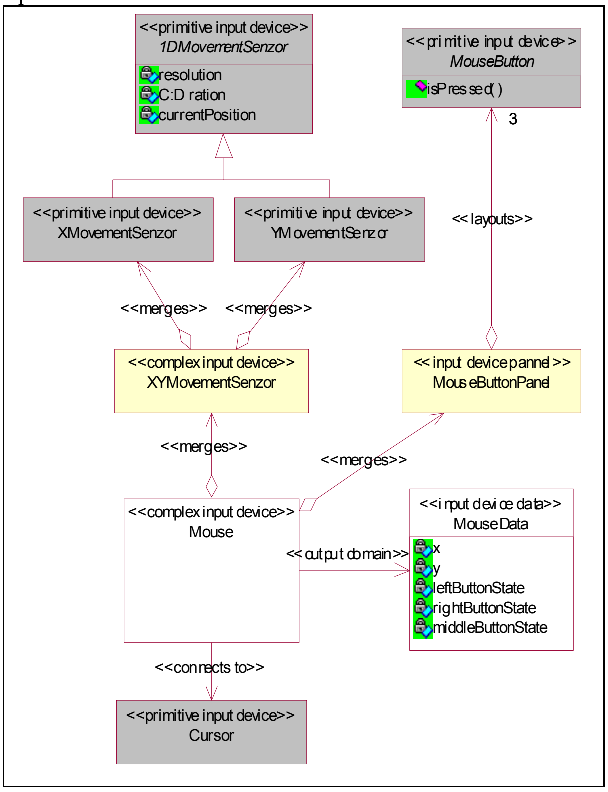

Figure 2: Using the input device ontology for describing a computer mouse

The ontology of user interface devices has two parts: the input device ontology and the output device ontology. In this ontology we have merged Card, Mackinlay and Robertson input device framework and Allanson sensory model in order to create unified framework that can be used for description of broader class of user devices (Card, 1991; Allanson, 2002). Our input device can be simple or complex. A complex device integrates one or more simple or complex devices. We have defined four types of integration mechanisms: merge, layout, connect and complex processing integration. A merge input device creates new input device merging two simpler. A layout input device layouts two or more devices on a common panel. A connect input device connects the output of one input device to the input of other input devices. A complex processing input device creates a new input device whose output is the result of some complex processing such as neural network processing or pattern recognition.

The output device ontology is similar to the input device ontology, so we will not discuss it here.

With these UI ontologies we have described UI devices used in our research including various electrophysiological input devices such as EEG electrodes. For example, Figure 2 shows the description of a mouse using concepts from our input device ontology.

4.3 The stimulus and environment ontology

In this ontology we have identified basic stimulus and environment concepts of interest for the HCI field. Main concepts of this ontology are the concepts of stimulus and of its sensible properties. For example, sound is a stimulus that has intensity (amplitude) and frequency (pitch) sensible properties, while light has wavelength (color) and brightness sensible properties. Using this model we have described various stimuli, such as light and sound, as well as their characteristics.

The environment part of this ontology describes various environmental conditions that can influence human interaction, including light characteristics, sound characteristics, temperature, air pressure and humidity.

4.4 Integrating ontologies: The ontology of multimodal interaction

The ontology of multimodal communication is our main ontology and it imports and connects concepts from the domain ontologies in order to allow definition of interaction on different levels. Figure 3 gives a simplified overview of concepts defined in this ontology.

The main concept in our ontology is the concept of a computing mode. We defined a computing mode as a form of interaction that was designed to engage some of human capabilities. User interfaces can be viewed as one-shot, higher-order messages sent from designers to users (Prates, 2000). While designing a user interface the designer defines an interactive language that determines which messages and levels will be included in the interaction. We have classified messages that a mode can send in three main categories: sensual, perceptual, and cognitive. These messages are connected with appropriate ontologies. A sensual message is a low-level stimulus effect aimed to excite some parts of human sensory apparatus. This effect can, for example, be used in EEG-based electrophysiological interfaces as a stimulus for evoked brain potentials. A perceptual message engages some of human perceptual skills such as pattern recognition or perception of three-dimensional cues. A cognitive message engages more complex human capabilities such as memory, attention or cognitive chunking.

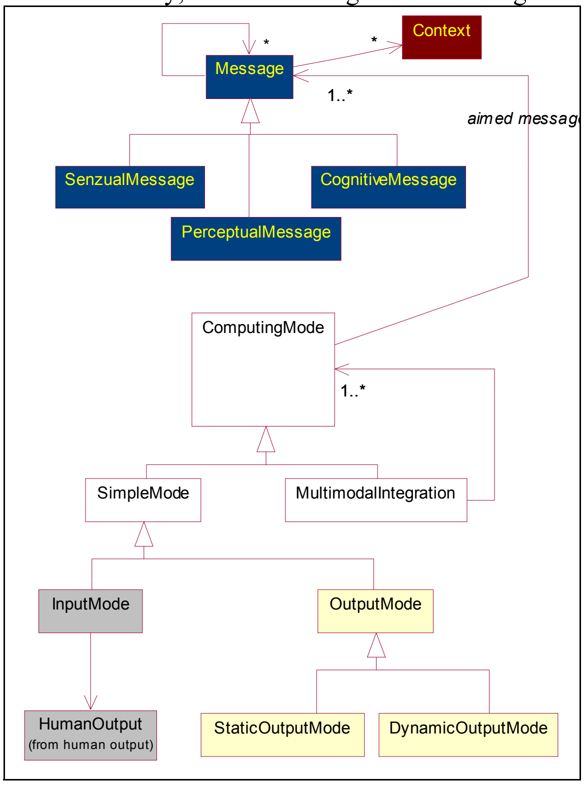

Figure 3. Structure of the multimodal interaction ontology represented as a UML class diagram.

A message is often connected with other messages. For example, all perceptual messages are a consequence of sensual effects. These relations among messages are important because in this way a designer can see what side-effects will be caused by his intention to send some message. In addition, each message is connected with contexts in which it works. A context defines the various conditions that a mode requires in order to produce desired effects. For example, linguistic presentation requires the user to speak the language in which the presentation is made. We have defined three main groups of the context: the environmental context, which defines environmental conditions, the user context, which defines user long-term and temporary characteristics, and the task context, which defines the context of current application task.

In our model, a computing mode can be simple or complex. A complex computing mode integrates other modes to create simultaneous use of various modalities, while a simple mode represents a primitive form of interaction. We have defined input and output types of simple computing mode. Input computing mode requires some user devices to transfer human output into a form suitable for computer processing. Output computing mode presents data to the user. A presentation can be static or dynamic.

The main advantage of this model is the flexibility that allows description of a user interface as a collection of modality messages at different levels of interaction (sensual, perceptual, cognitive). In addition, this model allows integrating of input and output modes in one integral unit.

5 Designing multimodal user interfaces using ontologies

In this section we will discuss various approaches to the design of multimodal user interfaces based on our ontology. Firstly, the unified ontology provides primitives for construction of high-level models and tools for multimodal interfaces. The additional benefit of using our ontology comes from the mapping of the ontology concepts into primitives available in existing user interface platforms.

5.1 The design of high-level models of multimodal interfaces: A case study

In our case study we will describe the environment for experimental evaluation of a multimodal feedback in dynamic pursuit-tracking tasks. The dynamic pursuit-tracking paradigm of interaction has valuable applications in fields of surgery, night vision or low visibility mission navigation, and flight navigation and orientation (Obrenovic, 2002). We have examined the effects of multimodal feedback in this task, extending the visual feedback with various sonification paradigms. The environment is implemented as virtual audio-visual scene using the Java3D package. We have used one graphics mode, and two different sound modes: variable position sound object representing target position in the space based on a stereo effect, and variable intensity sound object representing the distance from the target. The task of the user was to track moving object with the cursor as close as s/he can. The target object was positioned according to a predefined trajectory taken from a file, and coordinates of both the target object and the cursor were written to the result file for off-line processing.

With this environment we have conducted several experiments that have shown that supplementary acoustic presentation improves the quality of human-machine interaction and reduces errors during pursuit tracking. But only stereo sound showed statistically significant results (Obrenovic, 2002).

Figure 4 shows the high-level model of our multimodal pursuit-tracking environment. For creation of high-level models we have used UML extension mechanism. For each concept in our ontology we have developed appropriate stereotypes. We have developed scripts that export our model into several RDF files, each file connected with appropriate RDFS encoded ontology. These RDF files are later used in our tools together with ontologies RDF and RDFS files.

Multimodal pursuit tracking is a complex mode that integrates hand movement human output, and multimodal feedback. We used mouse as an input device. Multimodal feedback is a complex output mode that integrates a visual presentation, and two audio modes. A visual representation integrates static background presentation and animated target in order to attack user's visual motion detection perceptual mechanism. The first audio mode is designed to produce three-dimensional stereo effect using stereo cues available in the Java3D package, while the second audio mode warns a user about the error that he or she does by changing the intensity of sound.

Our ontology was very valuable in discussion of experimental results because we were able to connect our model with knowledge described in other domains. For example, the human perceptual ontology showed us that that motion paradigm used in our environment activates motion detection visual perceptual mechanism. The human cognitive mechanisms ontology showed that motion detection attracts attention, which means that our eyes will be focused on the moving object. Looking at the ontology of human anatomy and physiology we could see that the user will visually track the animation of the object in fovea and parafoeva eye sensor regions, and these regions are very sensitive to details. As a consequence, no less than 80% of the visual cortex is involved in processing of our animated target. But it means that the rest of the interaction space will be weakly covered with our vision. This explains why the benefits of using the stereo sound had statistical significance. The vision gave us a good sensitivity for details in local space, while stereo sound provides us with a context of the overall interaction space.

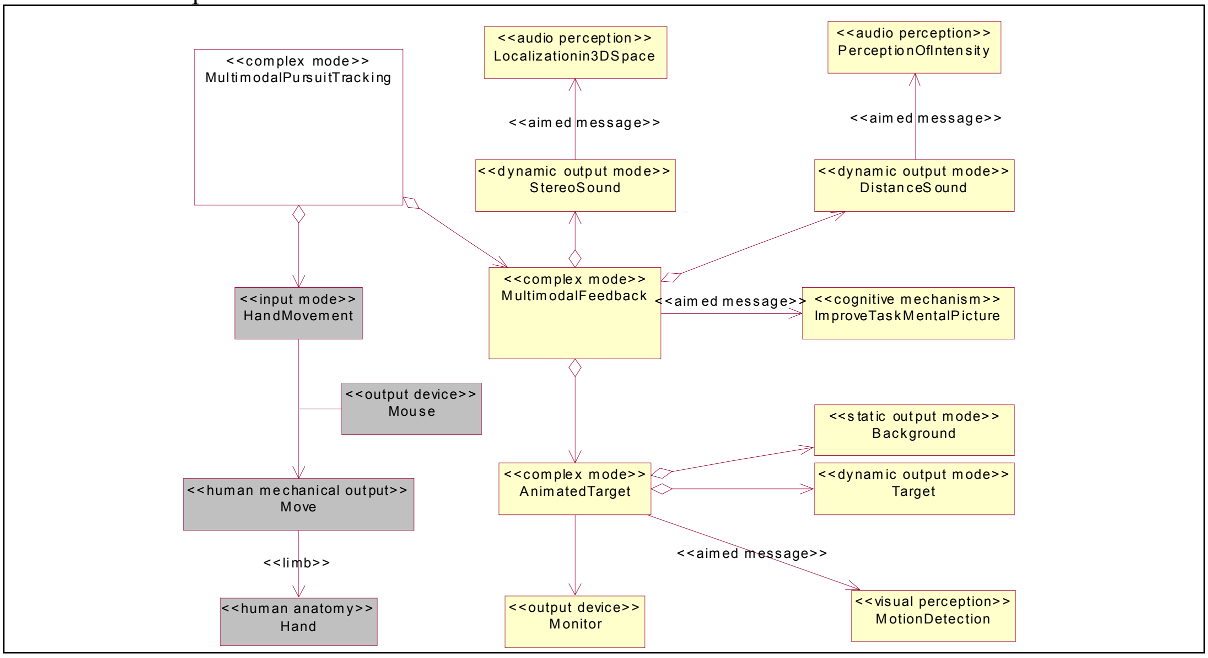

Figure 4: A model of multimodal pursuit tracking environment described with UML stereotypes defined from our unified ontology.

Also, we noticed that the perception of sound intensity is logarithmic, so our initial linear mapping of tracking error to the intensity of the distance sound was not the appropriate paradigm. Besides, we were able to see what muscle groups are active during user interaction, which explains some wrist fatigue reported by participants.

It is possible to use this model for automation of some phases of the design of new multimodal interfaces. For example, we have developed tools for generation of Java-based multimodal interface frameworks. These tools take as an input the RDF description of a high-level model, parse it and produce Java code files with abstract Java framework that represents a skeleton of designed multimodal interface. A designer can make use of generic mechanism supported in the framework, and extend this framework with platform-specific implementation of pre-specified modalities.

5.2 Mapping to UI Platforms

In order to make our unified ontology more practically feasible it is crucial to map the concepts from the unified ontology to primitives available in existing user interface platforms. For illustration, Table 1 gives simple examples of possible mappings of some high-level perceptual concepts into primitives available on textual platforms such as XHTML or text-to-speech platforms such as VoiceXML.

Effect	Textual platforms	Text-to-speech platforms
Pattern reco-gnition	Letter and word recognition	Phoneme, spelling and words
Group-ing	Similarity (font family) By proximity (words) Parallelism (rows of text) By proximity (paragraphs) Closure and surrounding (frames and tables)	Gender, age, name Word breaking rules Sentence breaking rules Silence
High-lighting	Size (bold, font size) Orientation (italic) Color (color of text) Flicker (blinking text)	Emphasizing Volume Rate and speed Pitch of a voice
3D cues	Interposition (z-index) Shadow	Available stereo formats

Table 1. One possible mapping of some perceptual effects to textual and text-to-speech platforms.

In addition to these simple effects, other platforms as X3D have more complex effects (such as motion parallax and 3D sound effects), which can be used to produce more complex cognitive effects.

With this mapping, it is possible to personalize the content since the user can choose the presentation form that he or she prefers. For example, alternative speech form of presentation can be very important for visually impaired users. In this way it is possible to evaluate existing user interfaces, such as numerous XHTML pages, since it is possible to see what will be the effect of the primitives used. This approach also allows transformation of content from one presentation format into another. For example, we have created a simple tool that parses XHTML and HTML files, creates new file with higher-level markup using terms such as group and highlighting, and which transforms this high-level markup into VoiceXML format for text-to-speech presentation.

6 Conclusion

In this paper we have presented an ontology-based approach to the design of multimodal user interfaces. The main result of our work is the unified ontology of multimodal interaction, where we have integrated the knowledge and common concepts from different domains of multimodal interaction in a uniform way. Using this ontology we have explored novel design approaches for multimodal interaction systems. We have also integrated our ontology with existing user interface platforms, which makes it possible to map high-level concepts from our ontology to one or more primitives from such a platform, or to evaluate existing user interfaces in order to see what perceptual or cognitive effects they produce.

The unified ontology and proposed design solutions may be very useful for designers and researchers of multimodal interfaces, as well as for lecturers and students of HCI courses. The unified ontology of multimodal interaction can provide them with a context of multimodal concepts where they could perceive many relations that are not always obvious, while our tools can help them create better multimodal interfaces. The ontology could also be used as a basis for collaborative creation of a broader set of knowledge about phenomena in the multimodal interaction field.

Our future work will include extension of the proposed ontology and inclusion of domains from other related communities such as user modeling and intelligent tutoring systems communities, that work with high-level user models. We are also working on the design of multimodal test environments, reusable multimodal components, and data mining tools for evaluation of various aspects of multimodal communication, which can exploit benefits of our unified ontology.

7 References

Allanson, J. (2002), "Electrophysiologically Interactive Computer Systems", *Computer*, IEEE CS Press, March2002, pp. 60-65.

Blattner, M.M. & Glinter, E.P. (1996), “Multimodal Integration”, *IEEE Multimedia*, Winter 1996, pp. 14-24.

Boff, K.R., Kaufman, L. & Thomas, J.P., (1986). *Handbook of Perception and Human Performance: Sensory Processes and Perception*. Vols. 1 and 2. Wiley, New York, N.Y.

Card, S.K., Mackinlay, J.D. & Robertson, G.G. (1991), “A Morphological Analysis of the Design Space of Input Devices”, *ACM Trans. Information Systems*, ACM Press, Vol. 9, No. 2, Pages 99-122.

Devedzic, V. (2002), “Understanding Ontological Engin– eering”, *Comm. of the ACM*, ACM Press, Vol. 45, No. 4, pp. 136-144.

Friedman-Noy, N. et al. (2001), "Creating Semantic Web Contents with Protégé-2000", *IEEE Intelligent Systems*, IEEE CS Press, Vol. 16, No. 2, pp. 60-71.

Jovanov, E., Starcevic, D., Marsh, A., Obrenovic, Z., Radivojevic V. & Samardzic, A. (1999), "Multi modal presentation in virtual telemedical environments", High-Performance Computing and Networking, *Lecture Notes In Computer Science*, 1593: 964-972.

Obrenovic, Z., Starcevic, D. & Jovanov (2002), E., “Experimental Evaluation of Multimodal Human Computer Interface for Tactical Audio Applications”, *Proc. IEEE Int'l Conf. Multimedia and Expo* (ICME 2002), Lausanne, Switzerland, Volume 2, pp. 29-32.

Philips, B.H., Rahman, M. & Järvinen, J. (2001), "Building a human factors ‘knowledge shelf’ as a collaborative information tool for designers", *Proc. Conf. Human Factors and Computing Systems - CHI 2001*, Seattle, Washington, pp. 98 – 103.

Prates, R., De souza, C. & Barbosa (2000), C., “A Method for Evaluating the Communicability of User Interfaces”, *Interactions*, ACM Press, Jan. – Feb. 2000, pp. 31-38.

Reeves, B. & Nass, C. (2000), “Perceptual user interfaces: perceptual bandwidth”, *Comm. of the ACM*, ACM Press, Vol. 43, No. 3, pp. 65 – 70.

Stanney, K.M., Mourant, R.R. & R.S. Kennedy (1998), “Human Factors Issues in Virtual Environments: A Review of the Literature”, *Presence*, MITPress, Vol. 7, No. 4, pp. 327-351.

Tan, D., Stefanucci, J., Proffitt D. & Pausch, R. (2001), " The Infocockpit: Providing Location and Place to Aid Human Memory", PUI 2001 Orlando, FL USA, ACM Press.