INTERNATIONAL DESIGN CONFERENCE - DESIGN 2004
Dubrovnik, May 18 - 21, 2004.

DESIGN 2004

AN ADVANCED VIRTUAL REALITY MULTIMODAL INTERFACE FOR DESIGN

R. De Amicis, G. Conti and G. Ucelli

Keywords: virtual reality (VR), multimodal interfaces, human computer interaction (HCI), computational linguistics

1. Introduction

The need for more natural ways of interaction with computers, together with the high degree of immersion provided by VR technology, have fostered enhancements in advanced graphical interfaces. This has led, in the last decade, to a constantly increasing interest in the field of Human-Computer Interfaces (HCI) and specifically, since the first developments during the 80s, towards human-computer Multimodal Interfaces. In this context the idiom *modality* is adopted to refer to the syntactic and semantic properties of a signal unlike the term "medium" which instead focuses on the production and transmission of signals [Cohen 1992]. In the real world people communicate with others using multiple sensorial channels such as speech, body gesture, gaze and facial expressions, and at the same way multimodal interfaces try to recreate this natural communication pattern as metaphor to interact with computers. Different modalities can be considered as complementary conceptual channels [Forbus 2001] that can transmit information not easily acquired spatially. Furthermore, such integrated approach is founded upon the effective support of human communication patterns [Oviatt 2000] that can provide, if combined, spatial description and mutual interrelation hardly achievable through other means. Researchers have proved that the most significant advantages of adopting multiple modalities is in the broader perceptual and conceptual bandwidth [Reeves 2000] available to the user to accomplish specific tasks and to interact with the computer. Several researchers [Reeves 2000] proved that different modalities can operate as complementary conceptual channels [Cohen 1992] conveying a greater spectrum of information on the spatial and semantic nature of the Virtual Environment (VR). Past researches by the authors [De Amicis 2002] [De Amicis 2003] have shown how the use of Virtual Reality (VR) and 3D input devices can significantly aid the initial design phases providing designers with tools that support their natural attitude towards creativity. The system presented pushes this approach further by introducing a Multimodal Interface for an immersive design tool. The advantages of using an immersive Multimodal Sketching system are manifold. First it becomes a valid aid for to the user's creativity and, second, it eases the use of the VR system since its interface tries to replicate the communication pattern typical of communication between humans (i.e. through concurrent use of gestures, gaze and voice). This allows both a decrease of user's fatigue, due to increased ergonomics within the workplace, and an enhancement of the efficiency of the interface since, as observed during its use, it reduces the need for interacting with menus and buttons.

2. Objectives

The system described in this paper, based on the author's previous work [De Amicis 2002] at the Fraunhofer IGD, pursues a user-centred paradigm during the initial stage of product design. This research effort has proved the benefits of the free-form sketching metaphor, a form of interaction embedded into the multimodal interface here described. Previous systems however forced users to interface with menus and icons placed on 3D panels which resembled traditional GUIs. The aim of this work is to further enhance the level of usability and effectiveness of the system by providing the means for the integration of multiple modalities, such as voice or gestures, into the design activity thus maximising the support for the user, bringing a positive impact on how design is performed. The designer in fact, rather than looking for menus and icon to press, can constantly focus on the design task asking for commands to be executed and pointing to elements to be modified. However the interface does not replaces the traditional GUI but it rather integrates with it so that, if required by environmental or working conditions, the user is free to use the traditional interface.

The system described is targeted to industrial design where fast development and rapid integration within the production cycle are crucial factors for competitiveness and commercial success. The project comprises the development of a state-of-the-art prototype that adopts a multimodal user interface based on the use of immersive visualisation technologies. This way the system effectively combines different modalities such gestures and speech to help operators accomplish design tasks in a natural user-centred way. The system can meet the requirements of the designer from the creative and artistic point of view, by providing flexibility, intuitive approach, fast prototyping, high quality of visual images, quick 3D verification. Finally the aim of the project is the implementation of a system according to a modular structure, by developing a number of different subsystems which can be activated/deactivated at wish. This approach yields a great level of flexibility and adaptability of the system to the various hardware or environmental requirements.

3. Related Work

Several researchers have tried to deliver intuitive yet efficient and accurate approaches to interaction with design systems and more generally to three dimensional environments. Since the first systems of the eighties a number of researchers have proved the efficiency of human-computer multimodal interfaces. Cognitive scientists have investigated how the design experience can benefit from the support of multi-sensorial, or multimodal, interactions [Reeves 2000]. Cohen et al. [Cohen 1999] have successfully integrated gesture and voice to interact with 3D environments. Their QuickSet system uses a handheld device to access a map of the scene which is used for simple interaction with the system while more articulated commands are expressed through voice. Several multimodal interfaces have been developed for a number of VE for a large set of applications such as medical [Burdea 1998], military training or simulation [Cohen 1999]. Chu et al. [Chu 1999] demonstrated the benefits that multimodal interactions can bring to VR environments when used for design purposes. The increase in usefulness and accuracy, measured through a number of tests, aimed at determining the efficiency of different combinations of modality during the performing of traditional CAD tasks. The research proved how this interface could speed up the required time by 5-10 times if compared to traditional CAD systems. The research highlighted the increase in performance when voice and gesture were adopted to perform spatial tasks or navigation related commands.

The suitability of adopting independent subsystems for multimodal application has been proved by several researchers [Cohen 1998] where commercial recognisers (e.g. speech, hand-writing recogniser etc.) were integrated into customised applications. In particular, researchers [Cohen 1999] have demonstrated how this approach can be extended by creating independent and intercommunicating agents for gesture, voice recognition or for multimodal integration.

4. The system architecture

The complexity required by the engineering and development of an efficient architecture for multimodal interaction has been tackled by adopting a modular software design. This way it has been possible to independently develop each subsystem according to a common communication protocol. Some modules are devoted to processing the different streams of input data and they provide the semantic-level interpretation for them. Other parts of the system instead provide the semantic fusion features. This way, the entire system can potentially be easily adapted to different requirements by simply developing new modules or modifying the functionalities of existing ones.

This architecture, from the logical point of view, can be represented in a layered fashion (see Figure 1). At the topmost level there is the interaction with the application itself, e.g. the design environment, which is based on the StudierStube library and which has been documented in previous works by the authors [De Amicis 2002] [De Amicis 2003]. The application is driven through a high-level interface, e.g. the form of interaction which complies with the metaphor adopted by the system. At the lowest level instead there is the set of devices adopted by the system which are used by the designer to interact with the system and which provide representation for the low-level interaction.

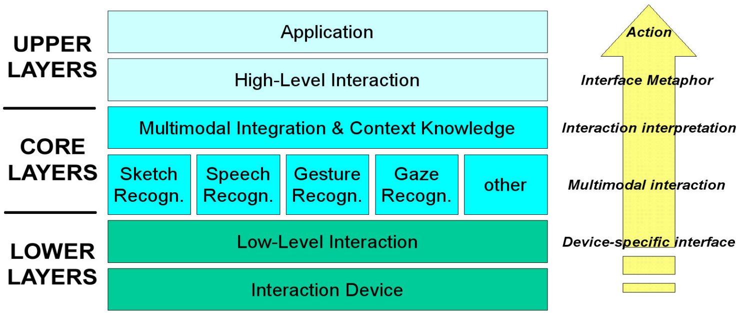

Figure 1. The architecture of the multimodal interface

The decoding of the information conveyed through the user's multimodal commands takes places in the intermediate layers. Here the information received by the devices is acquired, its semantic meaning is extracted and related to the knowledge that the system has of the state of the virtual environment. The first level of decoding takes place through the adoption of the OpenTracker library which allows for hardware independent abstraction of interaction devices. The system described in this paper has extended the functionality of the OpenTracker library by providing for instance full support for new hardware devices and for more advanced speech-driven functionalities.

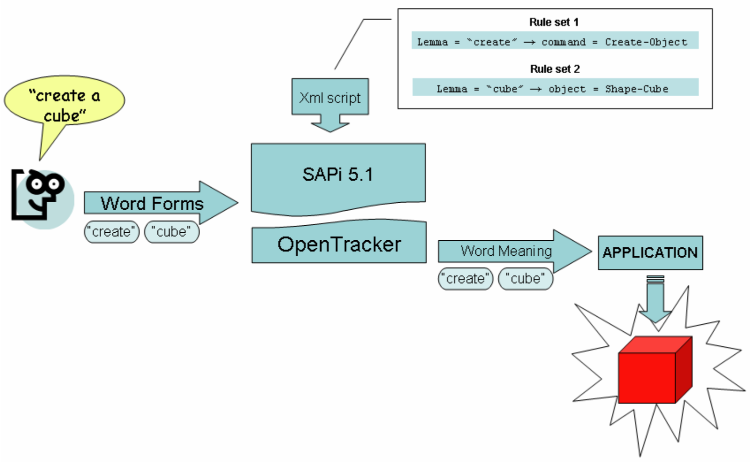

Figure 2. The system integration schema with SAPI and OpenTracker

A great deal of attention has been paid to the comprehension of verbal commands, give the major role these play if a natural interface is to be delivered. Specifically the system makes use of the speech recognition developed by Microsoft which is accessed via the Speech API (SAPI 5.1). The adoption of OpenTracker, which provides support for SAPI 5.1, has made it possible to assign to voice commands design actions. This way, for instance, the user can start the creation of a curve by simply speaking out the relevant command rather then pressing a button. The line then is drawn using the hardware available, i.e. a tracked pen or a virtual glove. The specification of the correspondence between spoken utterances and commands is handled via an Extensible Mark-up Language (XML) file which is read by the system and used by the SAPI engine to isolate the corresponding lemmas. The XML code, being text based, is human-readable thus easing the deployment phase and it is supported by a wide range of commercial software. Specifically, in order to increase the performance of speech detection, SAPI adopts a so-called Context-Free Grammar (CFG) which allows the modelling of the recogniser's language model. This information can be both statically declared through XML scripts and dynamically changed at runtime. This approach allows, if the application is carefully designed, to create very effective interfaces. Static grammars have been created for common commands, i.e. navigation, selection, whilst dynamic grammars that are activated, when required, for each specific status the system. For instance when the user declares the width of an object a *numerical grammar* is automatically activated and it translates the correct words into numerical entities. This approach helps the system to extract the semantics of the utterance relatively to the knowledge the system has of the current status (the word "**point**" as in "twelve **point** five centimetres" or "new control **point**") it reduces miscomprehensions and makes the process of recognition more efficient.

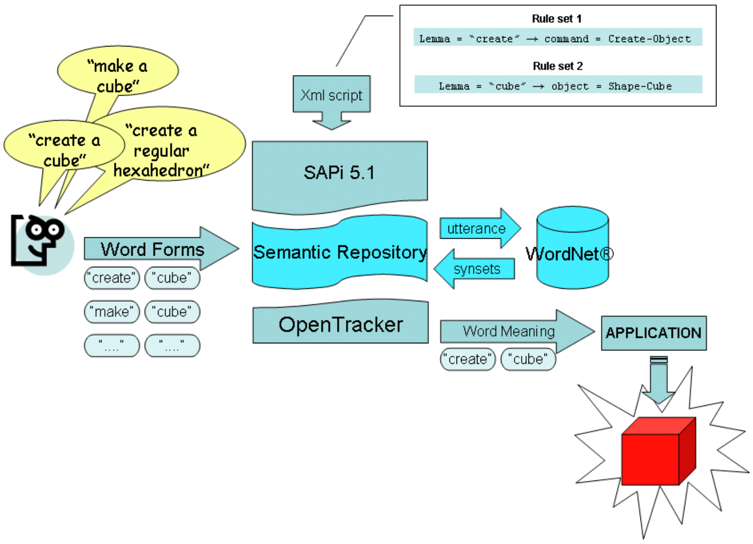

Figure 3. The system integration with WordNet

Using CFG it is possible to relate utterances with semantic information which in turn can be used by the system to activate relevant functions. For instance it is possible to link to the phrase "create a cube" a number of semantic information which specify: Action → "create" | Target →"a cube".

Whenever the user orders the systems to "create a cube" the semantic attributes are sent to the system and, in the previous example, the cube is created (see Figure 2). This approach, although powerful, it also makes the speech interface more rigid to deploy since it obliges the user to know the exact list of commands to be spoken. If the designer commands "make a cube" (rather than "create a cube") the system would not recognise the utterance and the command would not be activated. A solution to this approach is to provide SAPI with a more comprehensive CFG which lists utterances with same semantics activating a single command. This way it is possible for the programmer to hard-code a number of alternative commands that can be spoken by the designer to activate the same function. In the previous methodology this would be possible only by writing the relevant XML tags which make the "create" and "make" utterances start the same function e.g. the function creating the cube.

The advantages of using this approach are evident and manifold, from the increase in the usability of the system to the reduction of the overhead required to interact with complex three dimensional interfaces. However although this approach yields more flexible interfaces it is time-consuming, error-prone, not comprehensive and hard to maintain. To solve this problem the research team has endorsed a more linguistic-oriented strategy. The design environment, and in particular its multimodal subsystem, has been interfaced with WordNet. This is a lexical database where lemmas such as nouns, verbs, adjectives are catalogued according to their underlying lexical concept and can be retrieved using their semantic and lexical relations. Specifically, the system takes the lemmas specified by the user (e.g. "create" and "cube") and it performs a morphological analysis, e.g. identifying the list of possible inflectional endings relative to its syntactic category which can be removed from the word. Any eventual inflectional endings, i.e. third person verbs or plural nouns have the ending "s", is thus removed. The "normalised" word is then checked against the lexical database and, if available, identified as noun, verb etc. Finally the term is extracted together with the list of lexically related nouns. The list of terms, which in this case corresponds to a list of synonyms, is passed back to the SAPI and added automatically to the relevant CFG used by the speech engine as illustrated in Figure 3. Following this approach the system is capable of automatically recognise the meaning of a wider set of sentences without the need of tiresome and error-prone hard programming. In the previous example the user would be able to speak sentences such as "make a cube", "create a cube" or even "create a regular hexahedron" and the system would automatically recognise the semantics of the command and activate the relevant function. A further advantage of this approach is in the improvement of the system's performances by simply installing newer version of WordNet or by specifically developing tailor-made subsets.

5. Results

The system, although at a prototypal level, can be thought of as a design tool that allows the designer to freely sketch and manipulate 3D-shapes within an immersive Virtual Environment augmented through natural human pattern interactions. Initial testing have showed increased flexibility and most importantly the system's user-friendliness by supporting interactions closer to human pattern communication such as gestures and speech. According to the authors a major achievement of this work is its semantic-oriented approach, which allows for a more adaptable interface capable of comprehending more efficiently the semantics behind the user's command. Furthermore this approach could lead to the creation of a semantic repository to be used to add semantic information to the data structure of the models. This data structure can then be used, at a later search, for retrieval of semantic related information. This increases the efficiency of the systems and potentially increases the information that can be included within the designed object.

Finally the introduction of the technology described goes towards the improvement of the workplace's ergonomics by adopting more natural forms of interaction with computers. Finally it is worth highlighting that the user is not forced to use the full set of multimodal features offered by the system but he/she can, at any time, use the traditional menu-based interface if any environmental, working or personal conditions require doing so. However the adoption of multimodal interfaces for workers in the design area can help to remove stress caused by operating current systems in constrained environments. In fact the sources of common sickness caused by the design workplace are related to the use of workstations and they are essentially due to inadequacy of their interfaces which are cause of discomfort, distress and, ultimately, of a number of sicknesses. The prototype helps to increase the workplace ergonomics by preventing physical overstress during extended period of use by using speech and gestures. More generally, the overall broad range of benefits brought by the system include a new workflow in the styling process, easy communication of ideas and improved creativity through innovative tools.

6. Future Developments

Future development will include the extension of a semantic-based approach to the whole range of multimodal streams, including gestures, gaze, and sketch recognition. This way it would be possible to interpret more articulated semantics within gestures and automatically to attach their information to the model. Thanks to this approach the user not only is enabled to sketch 3D shapes using the free-form approach, but he/she can simultaneously express properties through their voice, gesture, gaze and alike. Likewise the user will be able to retrieve data, for instance from previous works, through semantic recognition of features. This approach, borrowed from computational linguistics, will be brought into the industrial design field to enhance the overall effectiveness of the system and to handle the properties embedded in the design in a more efficient way. Part of the ongoing research activity is the development of the terminology required for the mathematical and semantic description of designs in the phase of generation of curves and surfaces within a real industrial scenario.

As proved by previous researchers [Oviatt 2000] spatial application can greatly benefit from the adoption of multimodality. Dynamic context-aware weighting could further increase the benefit during the design process since it could bring advantages from the expressiveness of particular modes during specific contexts. This would yield better performances through general enhancement and stabilisation of the system's overall behaviour. Therefore through the adoption of this approach the Free-Form sub-system will provide the morphology of the information whilst the recognition sub-systems would provide different semantics and the means for further interactions. The result would be a seamless integrated system capable of creating and accessing complex information during the conceptual modelling stage in an intuitive manner resembling the human pattern communication.

Finally further enhancements will investigate the possibility to adopt an agent-based approach using frameworks such as The Open Agent Architecture™. In particular, the adoption of a distributed paradigm will be explored to permit a further level of flexibility through a system that could be possibly, but not necessarily, deployed across multiple computers. The modular approach would bring a number of advantages over the traditional monolithic architecture by promoting flexibility and by encouraging code reuse. Furthermore this approach is intrinsically less error-prone since it encloses functionalities within the boundaries of a framed subsystem, it discourages hard-coded communication between subsystems and it finally contributes towards the adoption of a "distributed", "pervasive" or "ubiquitous computing" approach through the adoption of a distributed architecture.

Acknowledgements

Part of this work has been done within the project InSIDe, Intelligent Styling system for Industrial Design, financed by the Provincia Autonoma di Trento. Special thanks go to Dr. Dieter Schmalstieg for providing StudierStube. The authors want to thank the Aim@Shape consortium for the support of the research activity.

References

Burdea, G., Patounakis, G., Popescu, V., Weiss, R. E.,"Virtual Reality Training for the Diagnosis of Prostate Cancer", Proc. of IEEE International Symposium on Virtual Reality and Applications (VRAIS'98). Atlanta, Georgia, USA, 1998, pp. 190–197.

Chu, C. P., Dani, T. H., Gadh, R., "Evaluation of a virtual reality interface for product shape design", IEEE Transactions, Vol 30, 1998, 629-643.

Cohen, P. R., "The Role of Natural Language in a Multimodal Interface" Proc. of UIST'92 : The Fifth Annual Symposium on User Interface Software and Technology and Technology. Monterey, USA, 1992, pp. 143-149.

Cohen, P. R., Johnston, M., McGee, D., Oviatt, S. L., Clow, J., Smith, I.,"The efficiency of multimodal interaction: a case study", Proc. of the International Conference on Spoken Language Processing, ICSLP'98. Sydney, Australia, 1998, pp. 249-252.

Cohen, P. R., McGee, D., Oviatt, S. L., Wu, L., Clow, J., King, R., Julier, S., Rosenblum, L., "Multimodal Interactions for 2D and 3D Environments", IEEE Computer Graphics and Applications, July/August 1999, pp. 10-13.

De Amicis, R., Bruno, F., Stork, A., Luchi, M. L., "The Eraser Pen: A New Interaction Paradigm for Curve Sketching", Proc. of the International Design Conference. Dubrovnik, Croatia, 2002, pp. 465 – 470.

De Amicis, R., Fiorentino, M., Stork, A., Monno, G., "3D-Tape Drawing in Virtual and Augmented Reality", Proc. of the XIII ADM - XV INGEGRAF International Conference. Naples, Italy, 2003, pp. 166.

Forbus, K. D., Ferguson, R. W., Usher, J. M., "Towards a Computational Model of Sketching", Proc. of the 6th International Conference on Intelligent User Interfaces. Santa Fe, New Mexico, USA, 2001, pp. 77-83.

Oviatt, S., Cohen, P., "Multimodal Interfaces that process what comes naturally", Communications of the ACM, Vol. 43 N. 3, 2000, pp. 45-54.

Reeves, B., Nass, C., "Perceptual user interfaces: perceptual bandwidth", Communications of the ACM, Vol. 43, N. 3, 2000, pp. 65-70.

Dr Raffaele De Amicis - Advanced Virtual Reality Multimodal Interface for Design
Fondazione Graphitech
Via F. Zeni, 8 Rovereto (TN), Italy
Telephone +39 0464 443450, Telefax +39 0464 443470
E-mail raffaele.de.amicis@graphitech.it