VRIO: A Speech Processing Unit for Virtual Reality and Real-World Scenarios - An Experience Report

D. Kranzlmüller[1], A. Ferscha[2], P. Heinzlreiter[1], M. Pitra[2], J. Volkert[1]

GUP[1] and IPI[2], Joh. Kepler University Linz
Altenbergerstr. 69, A-4040 Linz, Austria/Europe
kranzlmueller@gup.jku.at[1] | ferscha@soft.uni-linz.ac.at[2]

Abstract

Human Computer Interaction (HCI) summarizes research and engineering activities related to the communication between human beings and all sorts of "computerized" machines. Within this domain, substantial amount of work is dedicated to the idea of using the human voice as a natural interface for accessing computer systems. The VRIO speech processing unit represents one example of such an interface, where users control the machine via spoken commands. While the application of VRIO was originally intended for Virtual Reality (VR) environments only, a major redesign of VRIO's architecture allows its application to arbitrary scenarios, e.g. within ubiquitous and pervasive environments. This paper describes the revised architecture of VRIO as well as examples of its application in VR environments and for real-world scenarios.

1 Introduction

Comparable to everyday life, communication problems between the human user and the machine may have substantial impact on either of the two communication partners and/or the surrounding environment and are thus not desired. For this reason, many on-going research projects investigate new or improved ways of Human Computer Interaction (HCI). An example is the VRIO prototype, a combination of a software framework and commodity-of-the-shelf hardware, which provides a flexible user interface within arbitrary computing environments.

The original idea of VRIO started in Virtual Reality (VR) environments (Burdea & Coiffet, 1994), in particular the room-sized, 3-D projection-based CAVE Automatic Virtual Environment (Cruz-Neira et al, 1992). While the visual output of the CAVE provides interesting possibilities to experience computer-generated scenarios, the options for user input are often not satisfying. Sophisticated input devices such as the wand, a 3D-like mouse with 6 degrees of freedom (Ware, 1990), require a significant amount of training, while usage of traditional input devices such as a keyboard, is limited by the user's position, posture, and movement in the CAVE.

This example calls for a more human-centred interface design (Landay & Myers, 2001), with possibly natural or intuitive human-computer interaction through multimodal input and output (Ark & Selker, 1999). The deficit of human to computer communication compared to computer to human communication (Damper, 1993) is addressed by a series of ongoing projects in areas such as speech processing and computer vision (Wahlster, 2000).

In this context, the approach of VRIO was to replace or enable parts of the user interaction by voice input. The user, wearing a headset with a microphone, was able to control the VR scenario via spoken commands (Kranzlmüller, Reitinger, Hackl & Volkert, 2001). Due to the invariance of the VR environment, the speech processing was performed on a dedicated workstation located closely by the user.

The application of VRIO in Virtual Reality environments demonstrated some of the limitations of the system but also the feasibility of using speech processing for HCI. Based on this experience, a second prototype of VRIO has been developed. The software framework of VRIO was completely reworked to provide a better suited level of abstraction. The resulting client-server architecture enables interaction between arbitrary command clients - not only speech processing - and the server on the one side, and between the server and arbitrary actuators on the other side.

This paper provides an overview of VRIO's architecture and two examples of its application, one within Virtual Reality and one in the real-world. The next section describes the architecture of VRIO and its basic functionality, while example applications are presented in Section 3. A conclusion and an outlook on future work in this project summarizes the paper.

2 Architecture of VRIO

2.1 Overview

The architecture of VRIO resembles the traditional client-server approach. The user interface is provided as a preferably simple and almost invisible device, which receives the commands from the user. The input device transforms the commands in machine-readable form and forwards it via the input interface to the central server. The server analyzes the commands and generates one or more corresponding controls, which initiates the requested actions on connected actuator devices.

An overview of the architecture is provided in Figure 1, with possible command clients on the left side of the server and example actuators on the right side. Of course, it is possible that a command client may also provide actuator functionality.

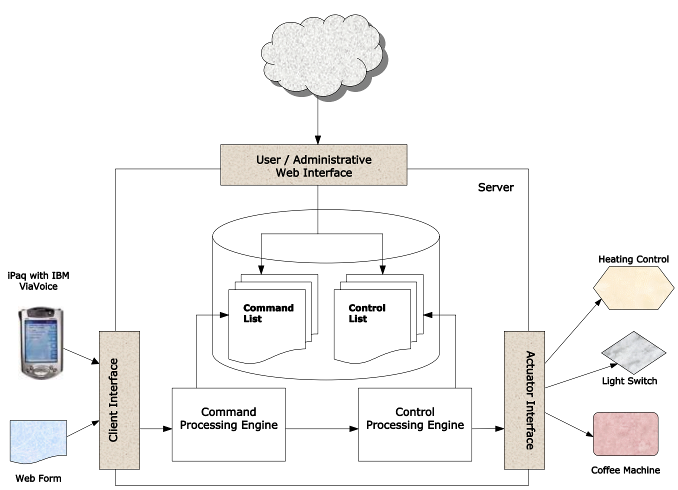

Figure 1: System architecture of VRIO

The communication between the client and the server, as well as between the server and the actuators relies on TCP/IP and HTTP, and can thus be easily integrated in many existing computing infrastructures. In fact, a variety of command clients and actuators, from simple web forms to self-made hardware instruments, have already been tested with VRIO. The mapping between commands and controls is provided by an XML request scheme, which can be easily adapted to different needs and scenarios. In particular, the system is able to dynamically adapt itself to different contexts, depending on the user's location and surroundings. The whole system is supported by an Application Programming Interface (API), which simplifies the usability of the framework for both clients and actuators. The API is (as much as possible) platform-independent in order to facilitate its usage on different devices, from embedded systems to sophisticated high-end installations such as the CAVE.

3 Example

As an practical example of this second generation VRIO, the hardware unit of the command client has been replaced by a personal digital assistant (PDA). Figures 2 and 3 display a VRIO user with a PDA, in this case a Compaq iPAQ. The PDA is equipped with a microphone for processing speech commands and a network interface card for communication with the server. The command client uses IBM's ViaVoice speech processing software (http://www.software.ibm.com/speech) to translate voice commands into corresponding XML requests. The requests are transferred over the network to the server, which maps the command to matching controls as indicated in Figure 1. The controls are then forwarded to the corresponding actuators in order to perform the desired activity.

3.1 Application of VRIO in the CAVE

According to our original intention, VRIO has been utilized within the CAVE Automatic Virtual Environment for different kinds of applications. Within the computational steering environment MoSt, the user controls the execution of a large scale high-performance computing application by navigating through a graphical representation of the program's states during its execution. The user is able to query the systems state, extract behavioural data, or modify parameters to change the program's behaviour.

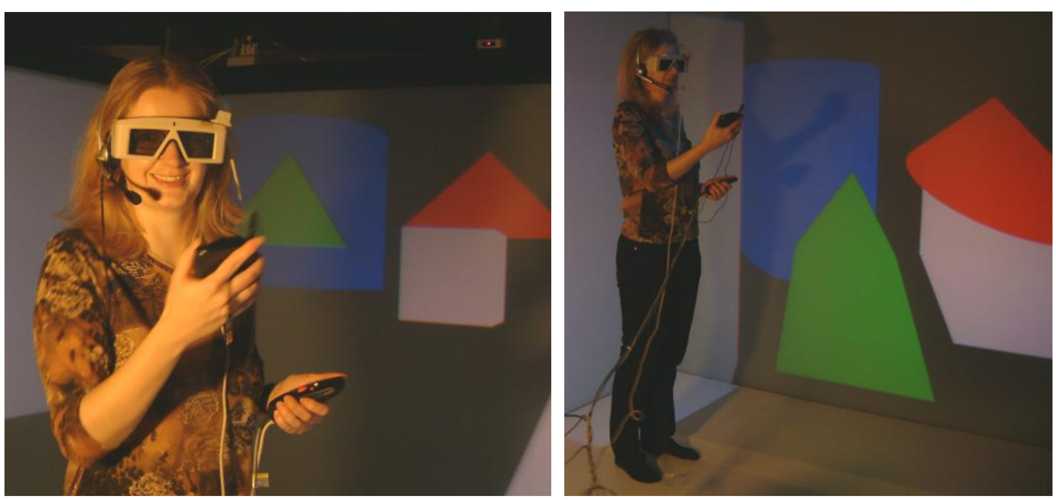

Figure 2: Application of VRIO with the CAVE Holodeck application

Another VR application of VRIO is the Holodeck 3D Editor. In this example, VRIO supports the user when generating arbitrary 3D VR scenarios, which can afterwards be used in artificial worlds. The 3D world is constructed interactively by placing and manipulating 3D objects. A variety of commands for handling 3D objects (e.g. cubes, spheres, cylinders, …) and other useful items (e.g. light sources) are provided. Objects can be generated, selected and moved within the virtual world. Their graphical representation can be modified by transformation such as scaling and rotation, as well as changing colours of the objects.
An example of the Holodeck 3D editor is given in Figure 2. The user is standing in the virtual world generated by the CAVE. Shutter glasses are required for providing the impression of 3D stereo pictures. The user's position is tracked through the glasses and the 3-D wand, a 6 degree-of-freedoms mouse, which is also used for object movement. The input client of VRIO (a Compaq iPAQ) in the right hand of the user is equipped with a headset to receive the spoken commands. Figure 2 contains some graphical elements of a simple scenario, which have already been positioned and manipulated by the user.

3.2 Application of VRIO in Real-World Scenarios

The redesign of VRIO increases its application domain from Virtual Reality to pervasive computing scenarios (Birnbaum, 1997). One example is a VRIO actuator for a standard interface card, which is used to connect arbitrary electric devices to a computer. With this approach, VRIO is able to switch and manipulate devices, e.g. to control arbitrary gadgets of consumer electronics. Another major effort of integrating this kind of natural user interface is being implemented in the Wireless Campus project, where students will be able to access administrative services of the Johannes Kepler University Linz. Related to this project is the public communication WebWall project developed at the IPI (Ferscha & Vogl, 2002).

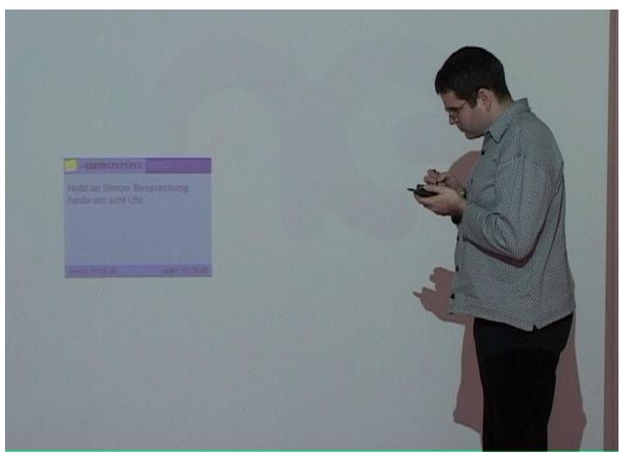

Figure 3: Application of VRIO in front of a public communication WebWall (Ferscha & Vogl, 2002)

The basic concept of the WebWall is based on the provision of large-scale displays at public places, which provides an interface to the World Wide Web (WWW), independent from the available input interface. Through WebWall's connection to the Internet and the cellular phone network, users can utilize the WebWall services as a kind of electronic pinboard for a variety of activities, e.g. placement of small advertisements, event notes, or even multimedia contents.

VRIO represents another sophisticated interface to the WebWall. Instead of typing the messages for the electronic pinboard, users can simply dictate their notes or activate arbitrary actions via spoken commands. An example is shown in Figure 3. The user stands in front of a WebWall, wearing VRIO (in form of a Compaq iPAQ) and speaking to the device. Immediately after the command has been submitted, a new window with the translated note opens on the webwall, containing the message of the user.

4 Conclusions and Future Work

The latest redesign of VRIO transformed the original Virtual Reality Input Output device from the self-contained world of the CAVE to ubiquitous computers and pervasive systems. The new and much more portable approach combined with the adaptable client-server architecture opens VRIO for a set of novel and interesting application areas. First examples in the CAVE and in the real-world have already demonstrated the feasibility of the approach. Several more applications are currently being developed.

An aspect, which has not been sufficiently covered so far, is (linguistic) context, which strongly influences the interaction between communicating human beings. Although the spoken words or gestures of humans are similar in different situations, the context of words or phrases enclosed by other words may trigger different "states" in the communication partners. Thus, the context itself must be considered as a primary source for input within any future human-computer interface.

Acknowledgments

Several colleagues at GUP and IPI have contributed to the development of VRIO. We are most grateful to Ingo Hackl, Bernhard Reitinger, Christoph Anthes, Edith Spiegl, and Simon Vogl.

References

Ark, W.S. and Selker, T. (1999). A Look at Human Interaction with Pervasive Computers. *IBM Systems Journal*, 38 (4) 504-507.

Birnbaum, J. (1997). Pervasive Information Systems. *Communications of the ACM*, 40 (2), 40-41.

Burdea, G., and Coiffet, P. (1994). Virtual Reality Technology. John Wiley & Sons.

Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V., and Hart, J.C. (1992). The CAVE: Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM*, 35 (6), 64–72.

Damper, R.I. (1993). Speech as an Interface Medium: How can it Best be Used?. Proc. Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers, Taylor and Francis, 59-71.

Ferscha A. and Vogl, S. (2002). Pervasive Web Access via Public Communication Walls, Proc. Pervasive 2002, International Conference on Pervasive Computing, Springer-Verlag, 84-97.

Kranzlmüller, D., Reitinger, B., Hackl, I., and Volkert, J. (2001). Voice Controlled Virtual Reality and Its Perspectives for Everyday Life. In: A. Bode, W. Karl (Eds.), Proc. APC 2001 - Arbeitsplatzcomputer, ITG Fachbericht, Vol. 168, Munich, Germany, 101-107.

Landay, J. and Myers, B.A. (2001). Sketching Interfaces: Toward More Human Interface Design. *IEEE Computer*, 34 (3) 56-64.

Wahlster, W. (2000). Pervasive Speech and Language Technology. In: Wilhelm, R. (Ed.), Informatics – 10 Years Back, 10 Years Ahead. Springer Verlag, LNCS, 2000, 274-293.

Ware, C. (1990). Using Hand Position for Virtual Object Placement. *The Visual Computer*, 6 (5), 245-253.