Using Hardware-based Voice Recognition to Interact with a Virtual Environment

Miguel A. Garcia-Ruiz[1] and Cesar R. Bustos-Mendoza[1]

Virtual Reality Laboratory, University of Colima, CEUPROMED, Colima, 28040, Mexico. Emails: mgarcia@ucol.mx, cbustos@ucol.mx

Abstract. This paper presents ongoing research on human-computer interaction in virtual environments using voice recognition systems. The paper starts describing voice recognition technologies (software and hardware based). The Virtual Reality Laboratory of the University of Colima has begun an initial development of a desktop virtual environment where a virtual molecule was manipulated and analyzed through voice commands. A hardware-based voice recognition system was used. Initial tests showed that the neural net, integrated in the voice recognition system, was fast and accurate enough to interact with the virtual environment. Further work is needed to add more voice commands to the system, and to extend experimental work, especially usability tests of the interface. This virtual environment system and voice recognition hardware is intended to use as a tool in biochemistry courses.

1 Introduction

One of the main purposes of virtual reality is to use more than one human sense to interact with information viewed in a 3D graphical environment [8], either to use it in educational, commercial or research applications [5]. One virtual reality specific attribute is to use special input devices to manipulate directly information contained in a graphical environment to facilitate its analysis and comprehension of information using, for example, a data glove or a 3D mouse [11]. According to [10], virtual reality is based on theories and rules described in the field of Human-Computer Interaction. Voice recognition techniques can be used to control information, for example, to activate states and get some results among other activities that happen in a virtual environment. Voice recognition systems can be hardware or software based. Commercial software-based voice recognition programs are Dragon's Naturally Speaking and IBM's Via Voice. These are used by an increasing number of users to ease text processing. Some of them are impaired users with learning disabilities [4].

Why is voice recognition useful in virtual environments? In the context of direct manipulation paradigm [11], it is possible that users keep their hands busy in any specific task by grasping a virtual object using data gloves and need to do another activity. They can use voice commands as an effective medium to perform the extra task and presumably increase the feeling of immersion in the virtual reality experience.

The hardware-based voice recognition systems are built with a circuit embedded with a neural network inside, including passive and active filters to drive the microphone, input and output ports, EEPROM memory among other components. The advantage of these type of systems are, among others, speed, accuracy, and lower costs than software-based voice recognition products. However, in general, the hardware based voice recognition systems manage less voice commands than software voice recognition systems do.

[2] carried out a comparative study about software and hardware-based voice recognition, and the best option selected was the hardware-based voice recognition technique. A summary of the comparative study can be seen in Table 1.

CHARACTERISTICS	HARDWARE-BASED VOICE RECOGNITION	SOFTWARE-BASED VOICE RECOGNITION
Speed	Fast (up to 1 second)	Depends on the computer
Accuracy	Very good (up to 95%)	Very good, but sometimes requires extensive training
Cost	Low (about U.S. $50)	Expensive (sometimes up to U.S. $1000)
Ease of install	Fair (needs to interface to a PC port)	Easy
PC Installation requirements	N/A, it can be connected to any PC. Just a PC port is needed.	High requirements (extensive RAM, a fast processor, considerable hard disk space)
Ease of programming	Easy	Fair. However, some commercial products do not offer APIs or programming libraries

Table 1. Comparison of technical aspects between hardware-based and software-based voice recognition.

2 System Design

Having seen the putative benefits of hardware-based voice recognition presented by [2], we used the VoiceDirect 364 speech recognition kit (showed in Figure 2). For our purpose of integrating voice recognition in a virtual environment, we used a virtual environments browser called DIVE (Distributed Interactive Virtual Environments), developed by the Swedish Institute of Computer Science [3] because this software is free for academic use, it can be used in various computer platforms, and it is easy to program, among other features. DIVE can

run Tcl programs and also has its own graphics language. A way to obtain the voice commands from the Voicedirect speech recognition kit is to connect it to a computer parallel port and read its output values (the n^{th} recognized word from the previously trained words) from the parallel port via a program. However, DIVE and Tcl languages do not control parallel port directly. Thus, a program made in Visual Basic was developed to read the parallel port and saved the data in a temporary text file. A DIVE/Tcl program easily read this file, and the data from the text file could serve to manipulate virtual objects in DIVE. Figure 1 depicts a schematic view of this process.

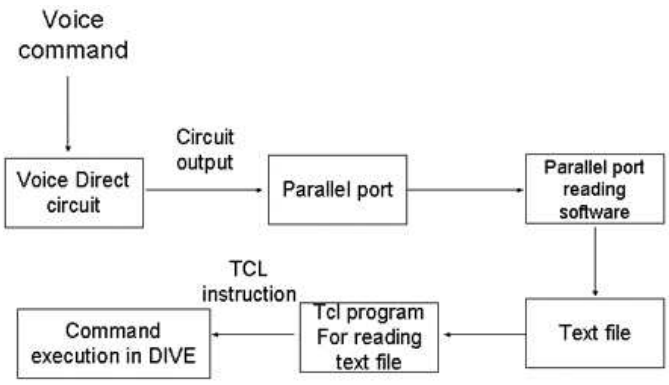

Fig. 1. Schematic diagram of the Voicedirect 364 circuit connection to the PC.

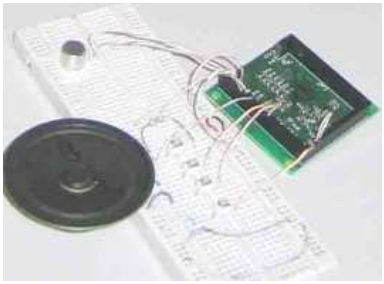

Fig. 2. Voicedirect 364 voice recognition kit.

A main problem we found was that Windows 2000, the operating system where the voice recognition system and the virtual environment system was implemented, does protect the computer data ports from being accessed by programs like the one we developed in Visual Basic to read the parallel port. To

solve this, we used a DLL software called Paraport (http://www.paraport.net/) to open the parallel port and then obtain the data from the Voicedirect circuit.

3 Implementation and System Functionality

One of the main objectives of the Virtual Reality Laboratory of the University of Colima is to offer a research environment for students. In it, an engineering thesis project was developed by [2] about using voice recognition technology to manipulate a graphical model of an amino acid.

[2] used a hardware-based voice recognition circuit, the Voicedirect 364, to control with voice commands three different molecular representations of amino acid Alanine, all placed in a virtual environment, and was visualized using program DIVE. Figure 3 shows the DIVE graphical interface. Molecular analysis was chosen because it has been successfully done with virtual reality technology, for educational applications [6]. The virtual environment had two main user tasks: one was about rotating the virtual molecule and the other was to change the molecular representation, using voice commands. To do this, three voice commands were used: "stick", "CPK" and "ball and stick". In addition, the molecule could be rotated in 4 directions using oral commands such as "up", "down", "left", "right". Figure 3 describes the molecular representations used in the thesis project of [2].

Because the Voicedirect circuit works as speaker dependent technology, the circuit was previously trained through voice commands using the microphone provided. Figure 2 shows the Voicedirect connections.

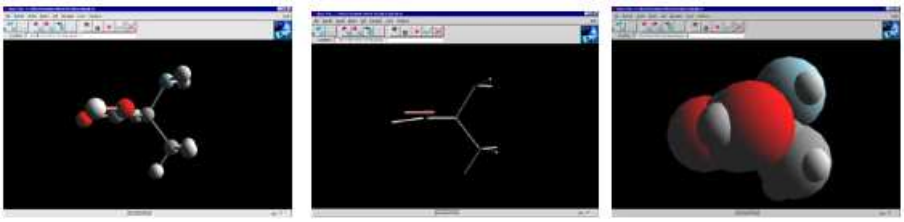

Fig. 3. The Alanine molecule representations ("ball and stick", "stick", and "CPK") that were selected using the Voicedirect 364 circuit.

4 Initial Tests

A pilot test was setup to try out the voice recognition system. Three Telematics Engineering students of the University of Colima were asked to test the system. First, they were explained the objective of the system and its functionality. The students were also instructed about how to train the circuit with the

voice commands, and the tasks (molecule rotation and change of representation, explained in the Implementation section). Preliminary results showed that the students could easily rotate the virtual molecule depicted in DIVE, and could change the molecular representations. However, background noise and the distance between the mouth and the microphone affected one of the tests done by a student. Since a small omni directional microphone was used, in future tests it will be necessary to use a headset microphone, to have it close to the user's mouth all the time. The voice recognition process was measured using a stopwatch, and lasted on average one second (from the voice command to the actual action of molecule rotation). It was a crude way to measure time, but in further tests we expect to have a program that can measure time on the millisecond scale.

5 Future Work

The next step is to apply usability evaluations to the system [1] [9]. It is also necessary to carry out accurate efficiency tests of the Voicedirect circuit. Since this virtual environment and voice recognition system is intended to use as a tool in further biochemistry courses, it is necessary to do pilot tests it in those curses with teachers and students, and do comparison studies with other types of input devices, such as a keyboard or a mouse to control virtual molecules.

6 Conclusion

This paper described a brief introduction of voice recognition systems and their possible use in virtual reality. In addition, it explained the design and development of a prototype of hardware-based voice recognition. A comparative study was carried out among software and hardware based voice recognition systems where it was found that the latter offers advantages relative to cost and recognition speed. A pilot test was carried out with the system, and preliminary results showed good accuracy and speed recognition. Further usability tests are needed to confirm this.

7 Acknowledgements

We would like to thank Adriana Ceja and Isabel Mendoza for their contributions to this research project. This project is funded by a PROMEP grant UCOL-EXB18.

References

1. Bekker, M.M., Vogten, L.L.M.: Usability of voice-controlled product interfaces. IPO Annual Progress Report **34** (1999) 111-124

2. Ceja-Castillo, A.E., Mendoza-Chavez, C.I.: Graphical analysis of amino acids using voice recognition techniques. Unpublished Telematics Engineering thesis. University of Colima, Mexico. School of Telematics (2003).
3. Carlsson C., Hagsan, O.: DIVE A platform for multi-user virtual environments. Computers and Graphics, **17** 1993.
4. de la Paz, S.: Composing via dictation and speech recognition systems: compensatory technology for students with learning disabilities. Learning Disability Quarterly **22** 173-182
5. Garcia-Ruiz, M.A.: Applications of virtual reality in education: a brief overview. Educacion 2001 **43** 37-40
6. Garcia-Ruiz, M.A.: Binding virtual molecules sounds good!: exploring a novel way to convey molecular bonding to students. Proceedings of E-learn 2002 Association for the Advancement of Computing in Education, Montreal, CA (2002) 16-19
7. Honeycutt, L.: Researching the use of voice recognition writing software. Computers and Composition **20** 77-95
8. Kalawsky, R.: The science of virtual reality and virtual environments. Addison-Wesley. Wokingham, UK (1993)
9. Nielsen, J., Mack, R. L.: Usability inspection methods. John Wiley and Sons. New York (1994)
10. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human-Computer Interaction. Addison-Wesley, Wokingham (1994)
11. Shneiderman, B.: Designing the user interface. Addison-Wesley. Reading, MA (1998)