Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 69 (2011) 3–8
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Developing speech input for virtual reality applications: A reality based interaction approach

Alex W. Stedmon*, Harshada Patel, Sarah C. Sharples, John R. Wilson

Virtual Reality Applications Research Team (VIRART), University of Nottingham, Nottingham, UK

Received 15 January 2010; received in revised form 20 September 2010; accepted 27 September 2010
Communicated by E. Motta
Available online 20 October 2010

Abstract

An input device should be natural and convenient for a user to transmit information to a computer, and should be designed from an understanding of the task to be performed and the interrelationship between the task and the device from the perspective of the user. In order to investigate the potential of speech input as a reality based interaction device, this paper presents the findings of a study that investigated speech input in a VR application. Two independent user trials were combined within the same experimental design to evaluate the commands that users employed when they used free speech in which they were not restricted to a specific vocabulary. The study also investigated when participants were told they were either talking to a machine (e.g. a speech recognition system) or instructing another person to complete a VR based task. Previous research has illustrated that when users are limited to a specific vocabulary, this can alter the interaction style employed. The findings from this research illustrate that the interaction style users employ are very different when they are told they are talking to a machine or another person. Using this knowledge, recommendations can be drawn for the development of speech input vocabularies for future VR applications.

Keywords: Virtual reality; Speech input; Human factors; Reality based interaction

1. Reality based interaction for speech interfaces

An input device should be natural and convenient for a user to transmit information to a computer (Jacob et al., 1993), and should be designed from an understanding of the task to be performed and the interrelationship between the task and the device from the perspective of the user (Jacob et al., 1994). The design of an input device should match the perceptual needs of the user and, as a result, the integration of input devices should follow a user needs analysis in order to map their expectations onto the attributes of the overall virtual reality (VR) system (Kalawsky, 1996). Building on a sound understanding of user needs, it is important, therefore, to analyse the task in the correct level of detail, so that the VR system and the VE that is developed supports user interaction and overall application effectiveness (Stedmon, 2003). A VR input device should account for the type of manipulations a user has to perform and be designed, so that it adheres to natural mappings in the way that the device is manipulated, as well as permit movements that coincide with a user's mental model of the type of movement required in a VE (Barfield et al., 1998). Despite such recommendations, there are very few established guidelines that detail what is required of an input device in terms of the parameters that pertain to a user and user performance within a VE (Kalawsky, 1996; Lee and Billinghurst, 2008).

With the development of reality based interfaces (RBIs) and new interaction styles that draw on users' knowledge, experience and expectations of the real-world, the aim is to develop human–computer interaction (HCI) metaphors in a digital world that are more intuitive and less constrained by technology (Jacob et al., 2008). With this in mind Jacob et al. (2008) present a framework for RBIs based on four fundamental themes:

- *Naïve physics*: people have common sense knowledge about the physical world.

*Corresponding author. Tel.: +44 115 9514068; fax: +44 115 8466771.
E-mail address: alex.stedmon@nottingham.ac.uk (A.W. Stedmon).

1071-5819/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2010.09.002

- *Body awareness and skills*: people have an awareness of their own physical bodies and possess skills for controlling and coordinating their bodies.
- *Environment awareness and skills*: people have a sense of their surroundings and possess skills for negotiating, manipulating and navigating within their environment.
- *Social awareness and skills*: people are generally aware of others in their environment and have skills for interacting with them.

The themes play a prominent role in emerging interaction styles as they offer universal paradigms for interaction which are not bound by cultural differences (Jacob et al., 2008). In relation to speech input, the RBI framework provides a clear structure for exploring both human–human interaction (HHI) and human–machine interaction (HMI). This paper presents an overview of speech input as a reality based interaction device before presenting the findings of a study that investigated reality based interaction of using speech in VR applications. The findings are then discussed in relation to the RBI framework along with recommendations for speech interfaces.

2. Speech input as a reality based interaction device

Speech is one of the most natural forms of HHI which most of us are able to employ in our daily lives from an early age (Stedmon, 2003). It is still not fully understood how we learn the subtle rules of syntax and grammar but what is clear, however, is that speech is a familiar, convenient, and spontaneous part of the capabilities people bring to the situation of interacting with machines (Lea, 1980).

In general, the use of speech for HHI requires little or no specific training except perhaps for tasks requiring specialized domain knowledge and vocabulary. When conducting HMI speech may offer advantages over keyboard skills or manual input, which are difficult without practise and training (Stedmon, 2005). Speech input presents considerable opportunities to harness reality based interactions; however, there is often still a requirement to learn task grammar and develop suitable mental models that account for how a machine might respond to spoken inputs. There is considerable progress in developing multimodal interaction, such as gesture, video tracking, and electromagnetic sensing, illustrating that as much as 97% of commands involve some degree of gesture based communication and that 94% of speech commands are preceded by gesture commands (Lee and Billinghurst, 2008).

2.1. Speech input for VR applications

Speech offers the potential to liberate users and allow a greater degree of freedom to interact within VEs as traditional input devices are often unsuited to simultaneous 3D navigation and control (Stanger, 1997). Speech allows hands free operation and could be particularly useful for navigation through easily accessible menus and short cut commands (Stedmon, et al., 2003). Furthermore, multi-user recognition systems support speech input in collaborative VEs allowing many users to interact with each other and the VE at the same time.

There is an increasing variety of input devices on the market that have been designed for VR use, ranging from tradition mouse and joystick devices, to wands, data gloves, haptic feedback devices, and speech interfaces (D'Cruz, 1999). Such variety may lead to individuals selecting an inappropriate input device that could compromise overall task effectiveness as well as impact on their overall satisfaction in using the VR application (Stedmon, 2003).

3. A reality based investigation of speech input for VR applications

In order to investigate the potential of speech input as an RBI device for emerging VR applications, two user trials and a comparative analysis designed to investigate issues associated with free speech were conducted. Two independent user trials were combined within the same experimental design to evaluate the commands that users employed when they were not restricted to a specific vocabulary. The study investigated speech input when participants were told they were either talking to a speech recognition system or instructing another person to complete a VR based task. Previous research has illustrated that when users are limited to a specific vocabulary and under time pressure, they are likely to revert back to prelearnt or more intuitive vocabularies (Baber et al., 1996) or when under increased cognitive load, they will take longer to recall the appropriate vocabulary (Stedmon and Baber, 1999). Furthermore, the match between the language users employ when using a speech input system and the language that the system can accept, or the habitability of a system, is a key issue in system usability (Hone and Baber, 2001).

In order to achieve free speech without the technical limitations of current speech recognition systems, a 'Wizard of Oz' paradigm was employed in a similar fashion to Baber and Stammers (1989) and Lee and Billinghurst (2008). Participants believed they were using a speech recognition system but, unknown to them at the time, an experimenter manipulated the VE based on their spoken instructions. In this way, by allowing users to employ any commands they wished, it was possible to investigate the following:

- The variety of legal and illegal commands that were used to complete the task;
- The style of speech used when participants were told they were talking to a speech recognition system;
- The style of speech used when participants were told they were talking to another person.