Correlations between Vocal Input and Visual Response Apparently Enhance Presence in a Virtual Environment

Christoph Groenegress, MRes.,[1] Mette Ramsgard Thomsen, Ph.D.,[2] and Mel Slater, D.Sc.[1,3,4]

Abstract

This work investigates novel alternative means of interaction in a virtual environment (VE). We analyze whether humans can remap established body functions to learn to interact with digital information in an environment that is cross-sensory by nature and uses vocal utterances in order to influence (abstract) virtual objects. We thus establish a correlation among learning, control of the interface, and the perceived sense of presence in the VE. The application enables intuitive interaction by mapping actions (the prosodic aspects of the human voice) to a certain response (i.e., visualization). A series of single-user and multiuser studies shows that users can gain control of the intuitive interface and learn to adapt to new and previously unseen tasks in VEs. Despite the abstract nature of the presented environment, presence scores were generally very high.

Introduction

THIS WORK PRESENTS an interface that uses some prosodic aspects of the human voice to manipulate abstract objects in an immersive virtual environment (VE). While human–computer interaction (HCI) based on speech has primarily focused on other aspects, such as speech recognition, most applications currently are limited by being applicable mostly within constrained environments.[1]

Aims and Hypotheses

In our work we examine to what extent humans can learn to use and control prosodic features of their voice as a means for interaction in a VE (H1). We also propose that a higher degree of interaction results in a higher sense of presence within the VE (H2). Learning the use of one's vocal apparatus for object manipulation in a VE is the central element in our study, while a secondary aspect is related to presence. Furthermore, we hypothesize that a shared VE based on this type of interaction can facilitate the construction of a perceived shared space between multiple users (H3.1), and that it can be used as a tool for extended human–human communication when both manipulate the same object (H3.2). H1 and H2 can be studied in an environment with 1 participant per experiment; H3 requires a multiuser scenario, in our case 2 participants per session.

Background

In virtual reality (VR) today, the most common devices for interaction are the wand and sometimes the dataglove.[2] They usually operate with 6 degrees of freedom. In more specific application domains (e.g. flight simulation, computer games), interaction devices are often modeled from real devices such as steering wheels; in some cases, interaction via 2D widgets in 3D virtual space may be appropriate,[3] and others employ hybrid approaches, using real objects to control the VE.[4]

Presence is commonly viewed as the illusion of being situated in a mediated environment rather than the actual physical location where the body is sited,[5,6] and it is concerned with understanding the psychological state or subjective perception that allows this to happen. While some theories are centered around action,[7,8] Slater and Usoh[9] introduced the term *body-centered interaction*, an idea based on the exploitation of appropriate movements of the whole body within an immersive VE likely to lead to greater presence—for example, by walking in place during locomotion.[10]

Method and Design

We conducted two group studies with the goals as stated in hypotheses H1 to H3. There was a single-user study with 8 participants (6 male, 2 female; mean age 34.5 ± 8.3 [*SD*]). In

[1]EVENT Lab, Facultat de Psicologia, Universitat de Barcelona, Spain.
[2]CITA, The Royal Danish Academy of Fine Arts, School of Architecture, Copenhagen, Denmark.
[3]ICREA - Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain.
[4]University College London, United Kingdom.

the multiuser study, we had 3 groups of 2 people, all of whom were male (mean age $= 29 \pm 4.4$). Both studies were carried out in a 4-sided CAVE-like system. Participants were equipped with a small high-quality clip-on microphone. We also implemented a simple tool for analyzing speech sounds. Regarding graphical output, we used nonuniform rational B-splines, or NURBS,[11] the parameters of which could be changed interactively based on the vocal input.

Procedure

Participants were recruited through advertisements posted across the university campus and paid 8 euros (about US$11) for participation upon giving informed consent. They were given the task of "using their voice to manipulate the shapes and find out what parts of their voice are important to affect the environment." There was no time limit, and at the end of each session, participants completed a modified version of a presence questionnaire.[12]

Environment

We chose an abstract and unusual environment to stress the necessity of action. Initially, it consisted of a single wave object slowly expanding horizontally in one direction. Each object responded to the user's voice. Volume and pitch would affect the wave's width and height respectively. Loud utterances thus resulted in a wider and fatter wave, while soft ones yielded a thin and narrow wave. Similarly, a high pitch elevated the wave, while a low pitch decreased it. Silence resulted in a "flat" wave and eventually led to its death, generating new one in a different position in space. Dead objects remained frozen in the space for some time. See Figures 1 and 2 for examples of wave behavior. During pilot runs, we also included further aspects, such as rhythm and voicing. However, more options for input only confused the participants, and no learning took place.

Results

Single-user study

A first glance, the questionnaire data reveals a surprisingly overall high degree of the reported presence score given the abstract environment (the mean score was 5.9 ± 0.6 out of a possible maximum of 7). Other scores relating to *subjective own achievement* and *control over environment* were signifi-
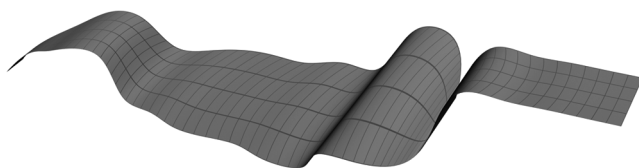

FIG. 1. All patterns are described from right to left, each segment is about 10 seconds long. A short period of silence is followed by a low-pitched soft tone, which briefly increases in pitch and then lowering again with a steadily decreasing volume.
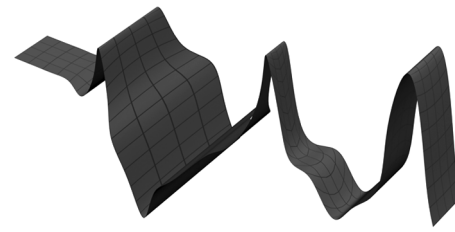

FIG. 2. The segment shows three peaks and valleys describing pitch. The volume about the last pitch is higher than for the two remaining peaks.

cantly lower (4.0 ± 1.7 and 3.8 ± 1.3 respectively). Age had a detrimental effect on these two scores, while the duration of the experience had a positive effect.

What is puzzling about the results is that despite low subjective control and achievement ratings and despite the abstract environment, presence scores were much higher than in comparable VEs with a more mundane and realistic scenario. The environment did not bear any relation to anything participants would expect to see or experience in the real world, so some other aspect must contribute to presence. There appears to be a quality beyond the conventional "sense of being there" that relates not to the realism of the environment but to whether it responds realistically to a person's own actions. We believe that the construct of presence is confounded and that participants respond to two different aspects of their experience. The first relates to the original idea of presence in regard to the illusion of being in a place.[13] The second refers to what we call plausibility, the illusion that the virtual events and situations perceived are really happening (even though the participant knows for sure that they are not). One of the factors that influences plausibility is when there are correlations between events in the virtual world and the actions of the participant. When presence is the only means by which participants in a questionnaire can respond, these two different aspects of the experience become inseparable: there is no other way for participants to report their experience than through the questions actually given in the questionnaire. In the particular case of this experiment, "being there" is not particularly relevant because there was no "there" to be in. However, the correlational aspect was important, since the participants could quickly learn that their actions were correlated with responses in the environment. Therefore, it is likely that the high presence questionnaire scores are not really to do with presence in the sense of being there but with plausibility in the sense of correlations between own actions and events in the world. These issues are discussed more completely in Slater.[14]

Multiuser study

We encountered unexpected social interactions among participants in the multiuser study, which led us to abandon it after three trials. Participants effectively entered into some sort of competition in each of the experiments. In one case, the participants attempted to outperform each other based on the volume of their voices. One of the participants, who had obviously lost the competition by that time, stood silently in one corner of the CAVE and watched the other trying to

interact. Another participant noted that "having two people made it more of a competition than an interaction."

Conclusions

We presented a novel way of interacting with a virtual environment by using some prosodic aspects of the human voice. During our experiments, we found that participants reported very high presence scores even though the VE was abstract and unusual. We believe that this did not relate to the sense of being there but to events in the environment being well correlated with user actions. A user study demonstrated that given sufficient time for training, humans can gain control of the interface and learn to adapt to new and previously unseen tasks in VEs (H1). Participants learned to use prosodic features of their voice (not involving speech) in a new way to control the environment. We also showed that there is indeed a correlation among potential for interaction (H1) and successful action and, in particular, presence (H2), or more likely, the plausibility concept referred to earlier. This strongly relates to work that was earlier called correlational presence,[15] which states that there is a direct relation between human action and the environment's response. However, we now think it inaccurate to include the word *presence* in this concept, since the plausibility we refer to is independent of the illusion of being in a place. Essentially, our brains must be very sensitive to correlations between our own activity and events in the environment.

The multiuser experiment was interesting because it unexpectedly brought into the picture a strong social element, competition between users, and therefore is simply not comparable to the single-user study. It raised many issues for further study. H3 was effectively left unassessed because of this issue. A future study could involve multiple participants interacting from remote physical locations.

Finally, there are some open questions regarding the benefits of using voice for interaction in this context. Throughout our experiments and pilot studies, we deliberately avoided using more than three voice parameters to manipulate an object and found that even three parameters produced by the same organ at the same time were too much to comprehend. In this sense, the voice may not be a very good medium to use for interaction.

Acknowledgments

Disclosure Statement

No competing financial interests exist.

References

1. Holmes JN, Holmes WJ, Holmes W. (2002) *Speech Synthesis and Recognition*. Bristol, PA: Taylor & Francis.
2. Burdea GC, Coiffet P. (2003) *Virtual Reality Technology*. 2nd ed. New York: Wiley.
3. Andujar C, Argelaguet F. Anisomorphic ray-casting manipulation for interacting with 2D GUIs. Computers & Graphics 2007; 31:15–25.
4. Lok B, Naik S, Whitton M, Brooks FP. (2003) Effects of handling real objects and avatar fidelity on cognitive task performance in virtual environments. Presence – Cambridge, Massachusetts 2003; 12:615–28.
5. Sanchez-Vives MV, Slater M. From presence to consciousness through virtual reality. Nature Reviews Neuroscience 2005; 6:332–9.
6. IJsselsteijn W. Towards a neuropsychological basis of presence. Annual Review of CyberTherapy and Telemedicine 2005; 3:25–30.
7. Zahorik P, Jenison RL. Presence as being-in-the-world. Presence: Teleoperators and Virtual Environments 1998; 7:78–89.
8. Flach J, Holden G. The reality of experience: Gibson's way. Presence: Teleoperators and Virtual Environments, 1998; 7: 90–5.
9. Slater M, Usoh M. (1994) Body centred interaction in immersive virtual environments. In Magnenat Thalmann N, Thalmann D, eds. *Artificial Life and Virtual Reality*. New York: Wiley, pp. 125–45.
10. Usoh M, Arthur K, Whitton MC, et al. Walking > walking-in-place > flying, in virtual environments. Computer Graphics 1999; 8:359–64.
11. Piegl L, Tiller W. (1995) *The NURBS Book*. London: Springer Verlag.
12. Slater M. Measuring presence: a response to the Witmer and Singer questionnaire. *Presence: Teleoperators and Virtual Environments* 1999; 8:560–6.
13. Held RM, Durlach NI. Telepresence. *Presence: Teleoperators and Virtual Environments* 1992; 1:109–12.
14. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc Lond (B)* 2009. In press.
15. Gillies M, Slater M. (2005). Non-verbal communication for correlational characters. In *The 8th Annual International Workshop on Presence: Presence 2005*. Cambridge, MA: MIT Press, 103–6.

Address correspondence to:
Christoph Groenegress
EVENT Lab
Facultat de Psicologia
Universitat de Barcelona
Passeig de la Vall d'Hebron 171
Campus de Mundet – Edifici Teatre
08035 Barcelona
Spain

E-mail: cgroenegress@ub.edu