Autonomous Identification of Virtual 3D Objects by Visually Impaired Users with Proprioception and Audio Feedback

Erico de Souza Veriscimo[(✉)] and João Luiz Bernardes Jr.

School of Arts Sciences and Humanities – EACH,
University of São Paulo, São Paulo, Brazil
{ericoveriscimo,jbernardes}@usp.br

Abstract. The World Health Organization estimates that there are 285 million people with severe visual impairment worldwide. With the advent of technology, 3D virtual environments are being increasingly used for several applications. Many of these applications, however, are not accessible to visually impaired users, creating a digital divide. Our goal is to develop a novel and low cost 3D interaction technique to allow the identification of virtual objects with autonomy using only proprioception and hearing. We developed a prototype implementing this technique and conducted preliminary tests with it, first with sighted users, only to test the prototype's basic functionalities and implementation, and then with a blind user with very promising results since the user was able to correctly and quickly identify virtual objects.

Keywords: 3D interaction · Accessibility · Visual impairment

1 Introduction

With technological advances, new devices for three-dimensional (3D) interaction are being created or becoming more available and less costly, contributing to the popularization of Virtual and Augmented Environments (VAEs). An example of this tendency is the extensive and growing use of Microsoft's Kinect, not only for entertainment but in several other applications, including research in several areas [1–4]. Many of these VAEs, however, are not accessible to visually impaired users, creating a digital barrier and excluding these users from certain activities [5]. The World Health Organization estimates that there are 285 million people with severe visual impairment worldwide [6].

Like any other citizen, those with visual impairment also have rights. The United Nations established in 1975 a declaration of rights specific to people with some form of disability [7] and these rights include:

- The inherent right to respect for their human dignity. Disabled persons have the same fundamental rights as their fellow-citizens, which implies first and foremost the right to enjoy a decent life, as normal and full as possible;
- Measures designed to enable them to become as self-reliant as possible;

- Right to education and other services which will enable them to develop their capabilities and skills to the maximum and will hasten the processes of their social integration or reintegration.

Sadly, there has not been much research done to improve accessibility to the visually impaired in virtual environments. There are works that attempt to minimize problems faced by visually impaired people, such as recognition of real objects using a camera (including that in a smartphone) as a sensor and augmenting the real environment with aural information about detected objects [7–9], for instance to help users when shopping and needing to identify products on a shelf. Section 2 of this paper discusses related work in more detail. This form of interaction can be replicated for virtual objects easily enough, but it removes autonomy from users, who are forced to depend on a third party (the system) to identify each object for them and provide information about it. Another issue with this automatic identification is that, if users do not already have a mental model of the real or virtual object, this form of interaction offers no aid in building one. If instead users are merely aided in identifying the object on their own they have greater autonomy and may being building a mental model of the object, which can be particularly helpful, for instance, for educational applications. While the sense of touch can be used for this with real objects, it is not so easily explored for virtual ones.

In this work, then, we propose, develop and evaluate a novel 3D interaction technique accessible to visually impaired users to allow the identification of virtual objects with autonomy, using only low cost, easily available and portable devices and exploring the senses or proprioception and hearing, without using the sense of touch, which would require some form of force feedback which, in turn, usually require more complex and expensive setups. The technique consists of allowing users to "touch" the virtual object with the tips of both index fingers, which are tracked in space using computer vision and registered in the same system as the virtual object, then giving 3D audio feedback when the object is touched, with a different tone for each finger, which allows users to trace the surface of the object with their fingers and use their position, sensed through proprioception, to build a mental model of the object and use this model to identify it autonomously. When one index finger or both are not near the virtual object (and are, therefore, not being tracked) a buzzing sound informs users so they can reposition the fingers. While initially the intention was to allow users to touch the object with all fingers and their palms, as they can do with real objects, preliminary tests showed that, without haptic feedback, this made identifying the point of contact with the virtual object very difficult, which is why only the tips of the index fingers may touch the object.

A prototype was built in order to investigate this technique, using the Leap Motion device to track hands and fingers and headphones for 3D audio feedback. Currently head position is not tracked by the prototype, so the system assumes the user is facing a specific direction to be able to provide proper directional sound cues, but adding head tracking would is a relatively simple improvement for future work. This prototype is briefly described in Sect. 3.

Before evaluating the system with visually impaired users, first the prototype was tested to make sure it functioned as specified and implemented the technique described

above correctly. Then a preliminary experiment was conducted with a blind user. These tests and their results are discussed in Sect. 4 which is followed by this paper's conclusion.

2 Related Work and Visual Impairment

According to a Systematic Review about 3D interactions accessible to visually impaired users, also by Veriscimo and Bernardes [11], the greatest concern for accessible interaction appears to be to aid in the task of navigation (exploring and moving within an environment), particularly in real or augmented environments but also in some purely virtual environments. 21.6 % of all papers included in the review dealt with navigation and only 3.8 % with object recognition.

These 3.8 % are composed of the already mentioned works of a Al-Khalifa [8] and Nanayakkara [9, 10]. They have a common concern of aiding visually disabled people to recognize objects using cameras or smartphone cameras, for instance products on shelves when shopping. The camera captures images of the object of interest, which are processed to attempt to recognize the object. If the recognition is successful, the system uses sound and names the product, vocalizing the word "shoe" for instance. We have already discussed how this lack of user autonomy in the process of object identification may bring problems to the user (not so severe in real environments, where the user could alleviate some of those problems by touching the physical objects, but much more so in purely virtual environments), including the lack of opportunity to acquire a mental model of the object. In that systematic review we were unable to find systems exploring 3D interaction to recognize virtual objects with user autonomy and particularly with low cost devices. The review summarizes the main applications and techniques for accessible 3D interaction in this context and which senses and devices these techniques explore but this discussion is beyond the scope of the present paper.

Instead, to develop our accessible technique to provide user autonomy in this task, we must better understand visually impaired users and how they interact with the world. When referring to visual impairment in this paper, we mean people that cannot see at all or those with severe difficulty in seeing. To interact with the environment, they use the following senses [12]:

- Hearing;
- Proprioception or kinesthesia (the sense of relative position of parts of one's own body and effort employed in movement [12]);
- Sense of touch;
- Sense of smell;
- Sense of taste.

Any interaction technique aiming to be accessible to visually impaired people must use only or primarily these senses, then. In the systematic review mentioned previously [11], the senses most often explored for interaction were, by far, hearing and then proprioception.

3   Prototype

Hearing and proprioception are not only very important senses used in interaction by visually impaired persons and were the most frequently explored in accessible 3D systems [11], they can also be explored with relative simple and low cost devices, such as headphones or speakers to provide aural feedback and sensors to track the position of user body parts. It was with that in mind that we developed the technique already described in the introduction, using the Leap Motion device [13] to track user hands and serve mostly as an input device and headphones for audio feedback. It is with the Leap Motion that we explore the sense of proprioception which allows the user to be aware of the positions of the two fingertips he "touches" the virtual object with (thanks to the registration provided by Leap Motion). Figure 1 illustrates our interaction technique, with the sphere representing a virtual object and the sound waves representing the audio feedback when each fingertip touches it.
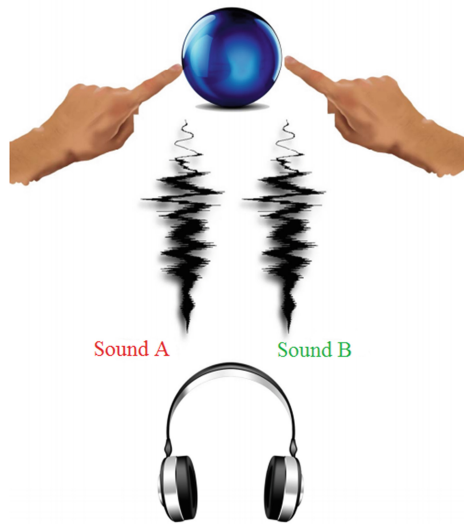

Fig. 1. Interaction Technique

This feedback uses 3D audio so the sound can originate in the point of contact, but since the two points are relatively close to each other for the size of the virtual objects we intend to use with this technique (limited also by Leap Motion's range of detection), the 3D position of the aural feedback is not always enough to identify which finger touched the object and we also need to use different tones for each. As mentioned before, we use a single fingertip from each hand to make it easier for the user to determine its position via proprioception, since earlier tests showed that using more fingers or the palm, as we had planned initially, makes this task much more complex, even though Leap Motion is capable of tracking all fingers, even when they occlude each other for brief periods of time during movement.

Because the user's fingers must be placed generally above the Leap device to be detected and because our target users cannot see the device's position, we use audio feedback to aid in that task as well. When sitting at a table they can either position the device themselves in front of them or be told generally where it is and the system emits a buzzing sound to inform the user if the sensor is unable to track one or both fingers for any reason (usually the hands being outside the sensor's field of detection) or if it is tracking more than just the two fingertips extended. The buzzing stops when tracking is properly resumed. Figure 2 shows some finger positions above the sensor illustrating one correct and three incorrect positions that would cause the system to emit the buzzing feedback.

Fig. 2. Finger positions relative to the sensor

To aid in debugging, in the conduction of the tests with visually impaired users and in some preliminary tests with sighted users the system also has the option to render on a computer screen the virtual object and two small spheres representing the positions of the fingertips of each hand, as shown in Fig. 3.
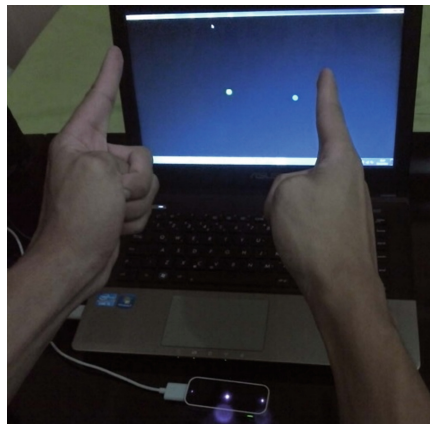
Fig. 3. Spheres showing fingertip positions

The prototype was developed in Java. The first version used JavaFx 3D rendering API [14], but this version of our system was discontinued because the API did not support 3D audio and we opted to continue development with a second version using the Java 3D API [15], which has native 3D sound support. This second version is the one that was used in the tests discussed in the present work.

Table 1. Participants

Participant	Gender	Age	Visual Impairment
A	Male	36	No
B	Female	61	No
C	Male	24	No
D	Female	19	No

4   Tests and Results

To test the prototypes basic functionalities before conducting experiments with visually impaired users, we first tested the system with four sighted participants. Since this was merely a test of basic functionalities and not of the interaction technique per se, we did not believe that the participation of sighted users to be a problem. Table 1 characterizes the participants in these first tests.

These initial tests were supposed to simply validate the following functionalities:

A. Whether the users can tell if their fingers are being correctly tracked or not and correct this situation so both are tracked. Users were blindfolded to simulate blindness and were asked to place their fingers over the sensor so both were detected correctly and keep them there for 30 s. The users were not shown the sensor or informed of its position and used only the buzzing feedback to accomplish this task, which was performed only once by each user.

B. Whether users can tell which finger touched the virtual object. As in the previous test, users were blindfolded (or the notebook was turned away from them) and were asked to explore the virtual environment moving their fingertips in 3D above the sensor until they could find an object. When that happened, they were asked which fingertip touched the object, based on the audio feedback with different tones for each fingertip. This time users were informed about the sensor's position before the test started. This task was repeated five times by each user with the virtual object in a different position each time, but the sensor in the same position.

Table 2 summarizes the results of these functional tests showing the time and success rate for each participant and each functionality. For functionality A, the time reported was the time until both fingers were detected correctly. The cumulative time shown for functionality B is the sum of the times necessary for the user to find the object with one fingertip all five times.

The results show that both functionalities are working properly in the prototype and that users could quickly find the sensor and virtual objects even when blindfolded. The time to find the sensor was between two and 4 s and averaged 2.75 s. Then, to find

Table 2. Functional test results

Participant	Functionality	Time	Success
A	A	3 s	100 %
B	A	4 s	100 %
C	A	2 s	100 %
D	A	2 s	100 %
A	B	15 s*	100 %
B	B	14 s*	80 %
C	B	13 s*	100 %
D	B	14 s*	80 %
*cumulative time			

the object, users averaged 2.8 s in the 20 trials. Out of those twenty trials, users could correctly identify which finger first touched the object 18 times.

Then we proceeded to conduct a preliminary experiment with a blind user. We were particularly worried at this point because, during the functional tests, sighted users reported difficulty in identifying or getting a mental model of the object they touched (which was a sphere), and even the authors had the same problem. This preliminary experiment with a visually disabled male participant happened in the following manner:

- The user was given two physical models of the virtual objects he should recognize, to make sure language would not be a barrier (the objects were a cube and a sphere, but the user referred to them as a rectangle and a circle during the experiment).
- He was then informed about how the technique worked, that he should place the tips of both index fingers above the area of the sensor, that it would stop buzzing when he did that and that a different tone would be played when each fingertip touched the object. He was also told that he could ask questions or abandon the experiment at any time. His participation was entirely and explicitly voluntary.
- He was not informed about the position of the sensor.
- There was no training, the first object presented to the user was already considered part of the experiment.
- The user had to differentiate between the virtual sphere and cube five times.
- The virtual object was selected randomly each time.

We were very pleasantly surprised with the results of this experiment. Out of the five times, with no previous training or even experience with the Leap Motion or 3D interaction in this way, the user got the object right four times. The only time he made a mistake was during the first interaction, perhaps due to lack of training and familiarity with the system. What happened in that instance was that, after finding the object with both fingers, the participant drew a shape with them crossing into the object instead of sticking only to its surface (and contact with the fingers was reported the whole time by the system), which was how he got the virtual object wrong. He was not informed that this is what happened, however, merely that he had gotten it wrong. This was enough for him to be more careful the next times and get all other objects right. Curiously, this

user interacted the system in a substantially different way compared to the blindfolded, sighted users. While they would touch the object once, then pull back and touch it again in other spots, he ran his fingertips in continuous motions over the surface of the virtual object. The task only took him a few seconds each time. At one point the user said with some excitement "I know it is a rectangle!", suggesting the system provided the autonomy we were aiming for (and the user was not informed that this was one of our goals). We are not sure about whether he referred to the objects as two-dimensional figures due to a lack of concern for mathematical formalism in his communication, lack of knowledge of the terms or whether he did build a two-dimensional model of the objects, since they were simple shapes with regular cross-sections and the user was indeed moving his fingertips mostly in a plane (but we believe this hypothesis of a two-dimensional mental model to be unlikely, since the user did have contact with 3D physical models in the beginning of the experiment). The experiment was recorded with the participant's permission.

5   Conclusion

The development of 3D interaction techniques accessible to visually disabled people is necessary to reduce the digital and social exclusion with the growing use of 3D virtual and augmented environments. It is desirable that these techniques be accessible regarding cost as well, taking advantage of devices with lower cost. Sadly, there is still little exploration of this topic in the literature.

We contributed to this area with a simple technique, using low cost devices, to allow visually impaired users to recognize virtual objects with autonomy, taking advantage of the senses of proprioception and hearing, and hope that it can be improved and extended or even that it can inspire more work in this area to reduce the barriers faced by visually disabled people in virtual environments.

The results of the preliminary experiments discussed in the previous section not only show we have a working prototype but also that the technique appears to have indeed enabled a blind user to successfully, easily and autonomously indentify virtual objects. These results appear promising enough to demand further research.

At the time this paper was written we had already submitted a formal experimental protocol to our institution's ethics committee, which was analyzed and approved, and we are currently conducting more experiments with blind users to verify the effectiveness of our technique with a larger sampling of participants. We also modified the system as a result of the preliminary tests, so that now it gives audio feedback only when the fingertips touch the surface of the object but not when they penetrate it past a small depth tolerance, to avoid the issue with users sinking their fingers in the objects and still hearing that feedback while no longer being able to trace the object's shape. These formal tests with more users and the modified technique and prototype are happening in the same general manner as described in Sect. 4 but with a third simple object added to the set.

After this experiment is over and its results analyzed and published, however, we will still have some important questions left to be answered with more experimentation. To check whether participants build 3D or 2D mental models of the objects while using

our technique, we plan to conduct an experiment asking them to identify different 3D objects that have the same cross-section, such as cylinders and spheres. We also intend to verify whether users can identify more complex 3D objects using this technique with a similar experiment. We very much would like to know how well users build this mental model of the virtual objects using our technique, but we are still having trouble designing an experiment to verify that. We also plan to build applications using this technique, the first of which will be an aid to teach 3D geometry for visually impaired students. Finally, it has been pointed out to us that this technique could be useful not only for visually disabled persons, but also to sighted users in situations where they cannot use sight to interact with 3D objects, be it because of darkness, for instance, or because they need to focus their gaze in some other direction but can interact with the virtual environment with their hands. We did notice a difference in how participants with and without sight used the system in our preliminary tests, inconclusive as they were due to the small number of participants, but we would also like to verify whether sighted users can be trained quickly to take advantage of this technique and how well they perform.

With all this we hope to add a small contribution in allowing visually impaired people to interact more with virtual objects and environments, be it for education, entertainment or professionally, to promote their inclusion in this context and greater social equality.

References

1. Lim, C.-K.: Application of kinect technology in the design of interactive products for chinese senior citizens. In: Stephanidis, C. (ed.) HCII 2013, Part I. CCIS, vol. 373, pp. 51–55. Springer, Heidelberg (2013)
2. Sousa, T., Cardoso, I., Parracho, J., Dias, P., Sousa Santos, B.: DETI-interact: interaction with large displays in public spaces using the kinect. In: Streitz, N., Markopoulos, P. (eds.) DAPI 2014. LNCS, vol. 8530, pp. 196–206. Springer, Heidelberg (2014)
3. Maike, V.R.M.L., Neto, L.S.B., Baranauskas, M.C.C., Goldenstein, S.K.: Seeing through the kinect: a survey on heuristics for building natural user interfaces environments in 8th international conference. In: UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014, Proceedings, Part I (2014)
4. Rodriguez Esquivel, J.C., Meneses Viveros, A., Perry, N.: Gestures for interaction between the software CATIA and the human via microsoft kinect. In: Stephanidis, C. (ed.) HCI 2014, Part I. CCIS, vol. 434, pp. 457–462. Springer, Heidelberg (2014)
5. White, G., Fitzpatrick, G., McAllister, G.: Toward accessible 3D virtual environments for the blind and visually impaired. In: Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts. DIMEA 2008, vol. 349, pp. 134–141. ACM, New York (2008)
6. WHO, World Health Organization.: Global data on visual impairments 2010. Geneva, 17 p 2010. Web. 21 Nov. 2016. http://www.who.int/entity/blindness/GLOBALDATAFINAL forweb.pdf
7. UN, United Nations. Declaração de Direitos das Pessoas Deficientes in: Assembléia Geral da Organização das Nações Unidas. 09 Dez (1975)

8. Al-Khalifa, A.S., Al-Khalifa, H.S.: Do-It-Yourself object identification using augmented reality for visually impaired people. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part II. LNCS, vol. 7383, pp. 560–565. Springer, Heidelberg (2012)
9. Nanayakkara, S., Shilkrot, R., Maes, P.: EyeRing: A finger-worn assistant. In: CHI 2012 Extended Abstracts on Human Factors in Computing Systems, 2012, pp. 1961–1966 (2012)
10. Nanayakkara S., Shilkrot, R.: EyeRing: a finger-worn input device for seamless interactions with our surroundings. In: AH 2013 Proceedings of the 4th Augmented Human International Conference (2013)
11. Veriscimo, E.S., Bernardes Júnior, J.L.: 3D Interaction accessible to visually impaired users: a systematic review in HCI International (2016)
12. Amiralian, M.L.T.M.: Compreendendo o cego. Uma visão psicanalítica da cegueira por meio de desenhos-estórias. São Paulo in Casa do Psicólogo, Fapesp (1997)
13. LEAP MOTION. Web. 06 Dez. 2014. https://www.leapmotion.com/
14. JAVAFX 3D. Web. 08 Dez. 2014. https://docs.oracle.com/javase/8/javafx/graphics-tutorial/javafx-3dgraphics.htm
15. JAVA 3D API. Web. 08 Dez. 2014, http://www.oracle.com/technetwork/java/javase/tech/index-jsp-138252.html