THE *LJUDSKOGEN*: A VIRTUAL REALITY REMAKE

Carlo Barone
KTH Royal Institute of Technology
cbarone@kth.se

Guilherme Diegoli Neto
KTH Royal Institute of Technology
diegoli@kth.se

ABSTRACT

We took inspiration from an audio-based installation to design a virtual installation using virtual reality software and experiment with sound interaction within this medium. While plenty of research exists on the topics of use of audio in virtual reality and use of body motion to modulate sounds, not much can be found about the combination of these two aspects, via the use of a head-mounted display paired with motion controllers. We designed a prototype in which the user can create their own installation and interact with it, and mapped most of the available interaction to sound modulations such as pitch and gain. Audio was also spatialized based on the position of the objects and the head. While we were not able to test the prototype extensively with users, we found that it enabled for unexpected combinations of different interactions, and that audio spatialization plays a big part in how this virtual environment can be perceived and interacted with. While the prototype might not be a substitute for a physical installation, it nevertheless highlighted the novel ways in which audio can be interacted with in VR.

1. INTRODUCTION

Virtual Reality (VR) technologies have enabled us to greatly enhance immersion and interactivity when it comes to software. While the first application that comes to mind when talking about VR is gaming, this kind of technology has also been proven useful in other areas, such as flight training [5], or rehabilitation [10]. However, one area that might not come up as often when exploring VR applications are museum installations. Some of the issues that could arise from this use of VR [9] are the difficulty curve of the installation and the fact that the installation should support potentially hundreds (if not thousands) of guests using it every day. This is a concern not only limited to VR installations, but also any installation which makes use of physical objects. These issues just tend to be exacerbated by the high costs and sensitivity of traditional VR hardware. Regardless, we believe that VR has a place in museum installations. Virtual environments (VEs) allows spectators to experience abstract environments which would not be possible to recreate with only physical means, at least not within a museum exhibition context. And with the use of motion controllers, these environments can also be interacted with in many ways, again some of which could be very difficult to create in the physical world.

For this project, we decided to take the idea behind the *Ljudskogen* ("The Sound Forest", in Swedish), a sound-based installation developed at KTH for Scenkonstmuseet by academics in Stockholm. Besides the angle of applying virtual reality to a museum exhibition concept, the following project has also been an exercise in how sounds can be placed and used in a VE, and how body motions can be mapped to sounds in order to enhance the already immersive interactions in VR.

2. BACKGROUND

Besides being an art installation, the original Ljudskogen [11] was also an exercise in designing a musical interaction model that was easy for guests to pick up and learn. The authors designed a string-based model, since, in their words, *"A string has clear affordances well known by most of the museum visitors and invites to different types of interaction such as plucking, bowing, punching, pulling, pushing, scraping and brushing"*. These strings were combined with sensors, lights and haptic floors to create an experience where guests could intuitively interact with the strings to produce simple musical sounds, either by themselves or with other guests.

We were inspired by the abstract nature and relative simplicity of Ljudskogen when coming up with this project. Due to it being a fixed installation, guests obviously are limited to how they can interact with the strings, so our main concept behind this project was to provide users with new types of interactions which only a virtual environment could allow for. In order to remain true to the original concept, however, we would have to make sure that not only should these interactions be simple and intuitive, but they should also provide some sort of auditory feedback to the user, which should create a playful environment for the user and potentially even engage the attention of the more music-inclined users.

Use of sounds in VEs is still an ongoing research, since the field of Virtual Reality has only recently been seeing widespread applications. In a case study, S. Serafin and G. Serafin [12] experimented with sound design to enhance immersion in a photorealistic VE, having found that while an appropriate soundscape can create a sense of place, listeners still couldn't fully recognize a place only by their sounds. The authors raise a series of points regarding sound design for VR, with the most relevant to this project being the consideration of head tracking and movement: proper spatialization relative to head position is crucial for immersion in VR.

The importance of 3D audio in a VR setting is explained in greater detail in Nasa's Technical Report "3-D Sound for Virtual Reality and Multimedia" [5]. Notably, 3D audio gives users an unparalleled situational awareness within a VE, since focus on sound sources can be switched at will without the need to move the head or eyes. Nasa's report claims that 3D audio is as important for sonification in VR as texture mapping is for visualization in VR. A more practical example of the application of sound spatialization in VR can be seen in Walker and Lindsay's experiment [4], where participants would have to navigate a virtual environment based solely on spatial sounds. Fortunately, most modern game engines are capable of spatializing sounds without the need for any additional implementations on the designer's end.

Another important aspect to consider for our project is the mapping motion to sounds. In a paper, Hermann [13] proposed different categorizations to interaction loops in an auditory display. Considering a common structure where data is translated to sound, interactions can occur at the sonification (interactive sonification) or data (navigation) levels. However, interaction can also occur as a means to provide data for the sonification system in real time; Hermann refers to this as "auditory biofeedback", which has users directly interacting with the sounds instead of simply controlling the parameters of an autonomous sonification system. Motion-controlled sonification fits this description, since the user's own motions are what provides the sound engine with its data.

Plenty of research exists on the topic: in a joint study between the universities of Victoria and Princeton [1], a framework was developed for the sonification of whole-body motion capture data. The framework was experimented with instrument-playing motions, motions based on emotion, and motion from individuals with Down Syndrome. In each of the three cases, a different approach had to be taken on which markers to consider, and how to map their motion data to audio transformations such as gain or pitch shift. Another study [6] developed a similar setup for generating sounds out of dance motion capture data. It is worth noting that both of these studies are over a decade old and predate the advent of modern VR technology, indicating that an interest in mapping motion to audio goes further than applications in VR.

A more advanced study [8] experimented with motion capture applied to music performance, where the performers themselves could modulate the sounds from their instruments with motion tracking not only from their own bodies, but also from the instruments themselves. The performers found the system to be unobtrusive when compared to more traditional modulation controls, allowing them to focus on the performance better.

While these studies can provide some insight in how to map body motion to audio, it is important to point out that they deal with a higher level of complexity than what is intended in this project: both studies used a motion capture system which allows for tracking of every body part and joint. While this is not unheard of in the VR field, most consumer hardware is only capable of tracking the head and hands. Unfortunately, not many studies are available when it comes to sound design within the limitation of VR motion controls.

Studies have also been done the other way around: a study by Caramiaux et al. [3] experimented with gestural description of both causal and non-causal sounds (that is, sounds with a clearly identifiable source and sounds with no clear source). Causal sounds were found to provoke gestures that would mimic actions that could cause them (such as crumpling a paper bag), while non-causal sounds evoked more metaphorical gestures. This could be an indication that, when designing interactions around sound, artificial sounds might map better to simple, abstract gestures, which happens to be our intention with this project.

3. CONCEPTUALIZATION

Since the concept of Ljudskogen is already well-defined [11], we wanted to create an environment and set of interactions that emulated the general look and feel of the original installation. The importance of sound spatialization in a VE [4, 5, 12] meant that the position of the sound sources had to be well-defined. We opted to have the sound emit from the top and bottom of the strings, like in the original installation. Ljudskogen allows for guests to physically interact with the strings by pulling, plucking, knocking, among other such interactions, so these had to be represented in some way. Additionally, we designed interactions that would not be possible in a physical installation, such as creating and removing strings, moving and rotating strings around the environment, changing the strings' color and base sounds, and lengthening/shortening the strings.

With these interactions in mind, we then had to design a sound model that encompassed all of those actions. In a paper by Hunt and Hermann [2], attention is brought to the aspects of interaction, and the potential meanings and uses of sounds during an interactive process. When a person interacts with an object, they expect feedback in some shape or form. Visual perception is typically the first step of interaction, but other modes of feedback such as sound are also important in getting confirmation that the interaction was carried out successfully, or otherwise in learning when something is wrong. Serafin and Serafin suggest that, in the context of virtual reality, having a physical sound model instead of just samples would be ideal for representing virtual interactions [12]. Since most of our interactions are motion-based, it becomes a matter of mapping these motions into sound transformations.

Starting with the simple interactions (create, remove and change strings), we did not feel the need to directly model any sounds after them, since they would already provide instant feedback by their very nature: creating a string also creates a new sound source, removing a string removes a sound source, and changing a string also changes what sound it plays. The more relevant interactions for sound design were the plucking/hitting, moving and scaling of the strings. Moving the strings would be as simple as "grabbing" it with the controller, with either one or two-handed grasps being possible. Holding a string with two hands would also allow the user to scale it by moving both hands closer or further apart, akin to a "pinch zoom" interaction. Plucking would be done with the controller's primary trigger, and the action would be abstracted by a virtual sphere that can initially be pulled from the string and then released, vibrating along its origin until it stops.

This leaves us with the motion-based interactions. Our intention when designing these sounds was to instigate a sense of discovery as the user interacts with the objects in the VE, and give them the means to set up a scene and play around in it as they desire. For the "moving" interaction, the velocity of each sound source attached to the string was chosen as the input for the sound system, meaning both translation and rotation would have an impact on the outgoing sounds. For "plucking", we opted to represent the physics of a vibrating string as a sphere that can be pulled from anywhere within the string and then released, bouncing around its origin point until it comes to a stop. For scaling, we simply take the string's length as a data value. To give the interaction some depth whilst keeping it simple to understand, we decided to allow the different interactions to be compounded (such as plucking a string that is being held and moving it around).

4. DEVELOPMENT

The application was developed using Unity 2019.2.6f1 64 bit, with the SteamVR plugin for VR functionalities. An Oculus Rift S head-mounted-display (HMD) and Oculus Touch controllers were used during most of the development process, but the application was also tested on an HTC Vive Pro HMD and Vive Wand controllers.

The user is able to walk around the VE freely, or use the controllers to teleport around the VE. Interaction is done via "grab-spheres" attached one to each controller in the virtual display. Once a grab-sphere is over an interactable object, the user can begin an interaction with the buttons and triggers on the controller.

On the sound designing side, the sound editing open source software Audacity was used to create three main sounds from which have a first implementation, with the aim of prosecuting with further and more specified sound creation platforms. Nevertheless being this just a first basis on which to work, it turned out to be working in an unexpectedly interactively various way, which gave our prototype a yet wide variety of sound interaction tools.

Audacity allows the user to create sound waves in form of sine, sawtooths and square waves. For comparing them, three samples were created, one for each type of available wave. Eventually, the sine wave was chosen, since it had a softer impact in audition, in contrast with the sharpness of the sawtooth and the square wave. Once chosen the most suitable type of wave, we proceeded to elaborate it for obtaining the first three samples. A basic sine wave at 440Hz (corresponding, on the music scale, to an A_4) and with amplitude A = 0.8 was the starting point. The three samples were created as follows:

1. **FIRST SAMPLE**: The basic sine wave was edited adding a reverb effect, with the Audacity's default parameters, whose the most relevant were the 50% of reverberance, the 50% of damping, a both wet and dry gain of -1 dBs and a pre-delay of 10 ms. The second editing step was a transposition, moving the basic frequency from 440Hz to 115Hz, obtaining so an Eb_3. Eventually, a low pass filtering, set at 1000Hz and a 6 dBs per octave roll-off, was applied.

   This obtained sound is low pitched and plays a good role as a background sound, since, combining the softness of the sine wave, the reverb effect and the lowness of the pitch it gives the impression to fill the environment, being neither too invasive, nor too hidden to the listener's ears.

2. **SECOND SAMPLE**: the second sound creation steps started with transposing the basic frequency of the sine wave to 262 Hz and modifying its amplitude in 0.6, obtaining thus a C_4, also known as *Middle C*, due to its central position on a piano keyboard. To this sine wave a similar one, with halved frequency - hence, one octave below - was overlaid, and a tremolo effect was added, in form of a sine wave in turn, with frequency equal to 131 Hz, 0 degrees of starting phase, the 40% of wet level. This sound was used for the grabbing interaction.

3. **THIRD SAMPLE**: the last sample used as a starting basis the second sound. A third sine wave was added over the first two, pitched one octave above, and played on regular intervals of 3.5 s, consisting in two seconds of sound and 1.5 second of silence, in order to give an oscillating appearance, but with the possibility of being looped, with the scope of being implemented in the plucking interaction, having thus a "periodic" appearance, but being adaptable to a loop without being recognized as a made-up track; that is, for instance, if the oscillation were implemented artificially, say, with a fixed duration of the file with decreasing duration of each repetition, if the length of the plucking-stimulated oscillation were to be longer than the sound track's length, ti would start over with an unnatural long repetition interval after a series of shortening intervals, no longer corresponding to the oscillation period.

The sounds were then added to the Unity scene and set to full 3D mode, meaning that sounds are completely modulated depending on the head's position within the VE. The sound is outputted through the HMD's own set of speakers, giving a very natural feel to the localization of sounds while in VR. For the other modulations, the Unity engine provides components to apply pitch shift, modify gain and even apply filters to the sounds.

We experimented with modulating pitch based on the velocity of the sound sources first, with higher velocities translating to higher pitches. An issue was quickly heard regarding very fast movements which could generate jarring sound modulations, so a cap was added to how much velocity could affect sound. While this experiment was a clear indication that the physical models were functional, we opted instead to have velocity affect sound gain instead, providing movement with a more subtle but still perceivable sound effect. Pitch was instead mapped to string scale (which were also capped), with shorter strings emitting lower-pitch sounds and longer strings emitting higher-pitch sounds. This allowed the user to more finely control the pitch of the audio through the "pinch zoom" interaction.

Each string object has an "idle" audio file that plays on loop for as long as the string is in the scene, and a "pluck" audio file that only plays when the "pluck string" interaction is performed. When a string is plucked, the audio starts at maximum volume and is linearly faded out until the sound source is completely silent. This linear fade out is the same used to generate the vibrating pattern of the plucked object, with further pulls vibrating for longer than shorter ones.

5. EVALUATION

While it would be ideal to test the application with naïve users in order to observe emerging behaviors (if any), we were unable to arrange any sessions with such users during the time assigned for this project. However, some interesting characteristics could still be observed while testing the application amongst ourselves: firstly, the visual aspects of the application alone were already sufficient to provoke unexpected interactions and flows when an user without previous VR experience was introduced to the application. While it could be attributed to the user's first time in VR, they were quick to understand how the environment could be interacted with and soon attempted to combine several of the available interactions in unexpected ways, some of which prompted corrections to be made to the application in order to accommodate for them.

Upon testing a prototype of the application with full audio implementation, the relevance of spatial audio not only became prominent, but it also enabled a new set of interactions not expected at first: the ability of the user to modify pitch, move strings around and move themselves around meant that different modulations of the idle audio would arrive at different intensities and from different directions, creating different soundscapes for different setups of the user and strings within the scene. However, testing with new users would still be a requirement in order for us to be able to properly observe and identify emerging behaviors and potentially make adaptations to the sound model.

6. CONCLUSIONS

The implementation of a sound installation in VR has shown some of the strengths of this relatively new medium. While VR might not be as intuitive as a physical installation, it is nevertheless more intuitive than traditional input methods such as mouse and keyboard. Where the medium appears to shine, however, is in its capacity to exhibit novel ways of both taking input from the user, and presenting output to them. The head tracking capabilities of modern HMD paired with stereo headphones allows designers to create intricate soundscapes that can be navigated naturally with head and body movement. The resulting interaction could be compared to the “Random Access Lattice” installation by Gerhard Eckel [7], in which a physical object is used to navigate an invisible, virtual soundscape.

The final prototype is not without its shortcomings, however. One important aspect of the original Ljudskogen was allowing for multiple guests to experience it simultaneously, something that is currently not possible in our application. There are also the logistics issues that would arise if this installation were to be present in a museum, as was mentioned earlier in this paper.

There are many ways in which this prototype could be expanded upon. It still requires extensive evaluation, and experimentation with new users could reveal potential flaws or strengths of its interaction system. Further experimentation could also be done with gamification of certain elements from the application, such as players having to use the spatialized audio to locate and select specific objects.

Ultimately, we believe that this project has touched upon some of the emerging possibilities when it comes to audio in a VR setting, by taking inspiration from an existing physical installation and using the medium to build upon its model. While we do not believe our prototype to be superior or a substitute of the original in any way, it should still hold its own as a demonstration of what motion controls paired with virtual environments mean for audio design.

7. LINKS

Should the reader be interested the prototype or its source code, it is publicly available on GitHub, under the MIT License. A demonstration video of the project is also available on Vimeo. All the links are hereby stated:

- *https://github.com/gdn002/SoundForestVR*
- *https://vimeo.com/398870172*.

8. REFERENCES

1. Ajay Kapur, George Tzanetakis, Naznin Virji-Babul, Ge Wang, Perry Cook. 2005. *A Framework for Sonification of VICON Motion Capture Data.* Proc. of the 8th Int. Conference on Digital Audio Effects (DAFX-05), Madrid, Spain, September 20-22, 2005

2. Andy Hunt, Thomas Hermann. 2004. *The Importance of Interaction in Sonification*. Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004

3. Baptiste Caramiaux, Frédéric Bevilacqua, Tommaso Bianco, Norbert Schnell, Olivier Houix, Patrick Susini. 2014. *The Role of Sound Source Perception in Gestural Sound Description. ACM Transactions on Applied Perception*. 11. 1-19. 10.1145/2536811.

4. Bruce Walker, Jeff Lindsay. 2003. *Effect of beacon sounds on navigation performance in a virtual reality environment.* Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, July 6-9, 2003

5. Durand R. Begault, Leonard J. Trejo. 2000. *3-D Sound for Virtual Reality and Multimedia.* NASA/TM-2000-209606, NAS 1.15:209606, IH-010

6. Frédéric Bevilacqua, Lisa Naugle, Isabel Valverde. 2001. *Virtual dance and music environment using motion capture.* Proc. Of the IEEE Multimedia Technology And Applications Conference (MTAC), Irvine, 2001

7. Gerhard Eckel. 2011. Random Access Lattice.

8. Gerhard Eckel, David Pirrò, Gerriet Sharma. 2009. *Motion-Enabled Live Electronics.*

9. Marcello Carrozzino, Massimo Bergamasco. 2010. *Beyond virtual museums: Experiencing immersive virtual reality in real museums.* Journal of Cultural Heritage. 11 - 4. 452-458. 10.1016/j.culher.2010.04.001.

10. Maria T. Schultheis, Albert A. Rizzo. 2001. *The application of virtual reality technology in rehabilitation.* Rehabilitation Psychology. 46. 296-311. 10.1037/0090-5550.46.3.296.

11. Roberto Bresin, Ludvig Elblaus, Emma Frid, Federico Favero, Lars Annersten, et al. 2016. *Sound Forest/Ljudskogen: A large-scale string-based interactive musical instrument.* Sound and Music Computing 2016. 79-84.

12. Stefania Serafin, Giovanni Serafin. 2004. *Sound Design to Enhance Presence in Photorealistic Virtual Reality.* Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004

13. Thomas Hermann. 2008. *Taxonomy and Definitions for Sonification and Auditory Display.* Proceedings of the 14th International Conference on Auditory Display, Paris, France, June 24-27, 2008