From Here to There

Troy Lapensee
tlapensee@ufl.edu
University of Florida
United States

Lauren Newman
lauren.newman@ufl.edu
University of Florida
United States

Michael Pérez
michaelperez012@ufl.edu
University of Florida
United States

1 INTRODUCTION

The target population is children. Due to the COVID-19 pandemic and economic recession, many children have not had the opportunity to travel and experience environments like farms, jungles, and beaches in person. Many children are growing up isolated, without knowledge of what such natural environments are like. The interface will give children the ability to move through several environments and virtually experience animal sounds. The project will give children valuable experiences through a fun and interactive interface.

There are numerous existing solutions to this same problem. Certain amusement park rides have a similar concept; for example, the "Avatar Flight of Passage" ride at Disney World takes the user through several areas using spatial sound [10]. Amazon has an app called 3D Animal Sounds, which allows the user to look at models of animals and hear the sounds they produce [4]. Zoos offer the experience of seeing and hearing live animals in person [2]. Virtual reality experiences in jungles and other exotic environments exist as well [11].

There are several challenges associated with these existing solutions. Spatial sound experiences in amusement park rides require high costs when considering park tickets and travel. Zoos also commonly have high admission prices and usually require a substantial drive to get there. The project will be free to use and will not require any travel because children will be able to use the interface on any computer. The 3D Animal Sounds app is relatively similar to this application, except it does not allow the user to navigate through multiple environments. Children will be able to navigate between the farm, jungle, and beach in this project. Virtual Reality experiences typically and require the user to wear headphones and a VR headset, which can be expensive. The interface will only require headphones; the visual interface will be displayed on the computer's display.

The proposed solution is a 2D visual and audio experience for children. This experience will allow users to move a character in 2D through environments (farm, jungle, and beach) that will render 3D audio of animals and other non-speech sounds. The character will start at a farm, then will be able to move through a jungle, then through a beach. The art style will be a simplistic pixel art style. As the users travel through the scene, there will be sounds that come from the animals at their locations relative to the user's location. The features include a colorful visual experience that gives users some interaction along with spatial audio to accompany the visual experience. The users should feel as if they were actually walking through the environments in person. Success will be achieved if Troy's children and Michael's nieces (all under age 7) enjoy the auditory and visual experience and learn about animal sounds while using the interface. After finalizing the project, the children will use it and then estimate how much they enjoyed it from 1-10 (10 being the most enjoyable). The average of these ratings will be the metric used to evaluate the project's success.

Some challenges with this project are anticipated. Aligning the sound with the visual interface will be difficult. If the visual interface is not correct, it can take away from the user experience as a whole and more specifically can make the auditory experience less meaningful. To overcome this, the 2D environment will be designed with pixel type art, to make the focus more on the sound and less about the complexity of the graphics. Since the scene will be 2D, rendering spatial sound that is above the user, like the sound of a bird that has flown overhead, will also be challenging. Sounds for animals may simple have to be rendered on the same plane as the character, and let the type of animal sound indicate that they are flying overhead. For example, perhaps when the character passes a certain point, a bird is triggered to fly across the plane, but since it is a bird, the user should understand that it is overhead. This will need some testing to verify. Blending the animal sounds is anticipated to be challenging. If two animal sounds occur at the same time, it may be difficult to distinguish which animal is which. It is hypothesized that the timing of the sounds will make a significant difference in such a situation. If the sounds only slightly overlap, the user will have time to hear each animal noise. Figuring out the best timing for the animals so they can all be heard well at different positions, but such that the sounds do not overlap too much and give some sense of realism to the virtual environment, will be another obstacle. This can be overcome through extensive testing, listening carefully to the auditory experience, and tweaking the animal sound time intervals. Well built and tested graphics libraries that are easy to work with will be utilized and complexity of the visual presentations will be reduced, so that more effort can be dedicated to the audio portion of the project. The visual functionality should take much less time based on the implementation decisions.

2 METHODS

2.1 Participants

"From Here to There" will be a whimsical game geared toward younger children. The primary target audience will be kids ages 3-6 because they will be most amused and receptive to the fun, animal theme of the game. The relative simplicity of the controls will make it easy for young children to learn quickly and easily navigate the map, whereas an older child may get bored of it more quickly. The application will be computer-based, so computer access will be a necessary requirement for a child and their parent to participate in the game. Since the parents will be the necessary facilitators for the child to use the game, they are the secondary target audience, specifically younger parents with a basic understanding of computers.

2.2 Materials

2.2.1 Stimuli. The game will consist of three zones the user can move around in: farm, jungle, and beach. Each zone will have a different background sound playing that reflects the theme. This component was added because it will make the experience more immersive and pleasant as compared to only including animal sounds. To compliment the background noise, the 3D audio will be implemented with the various animal sounds. For instance, if the character is in the farm zone with a cow on their left, the user will hear cow sounds coming from that direction. As the user moves forward or backward, the perceived location of the sound will move accordingly. Each animal sound will be played on a timer such that each animal's noise is played intermittently in a random pattern, like they would in real life. This was decided because it avoids the over-stimulation and annoyance that can come from the multiple animals' sounds overlapping [9]. It also makes the sound more natural since animals don't normally make noise consistently or with a predictable pattern. It was initially considered to add a narrative audio component to the game, but then it was concluded that this may detract from the game's immersiveness and make it more difficult to follow for the targeted age group [8].

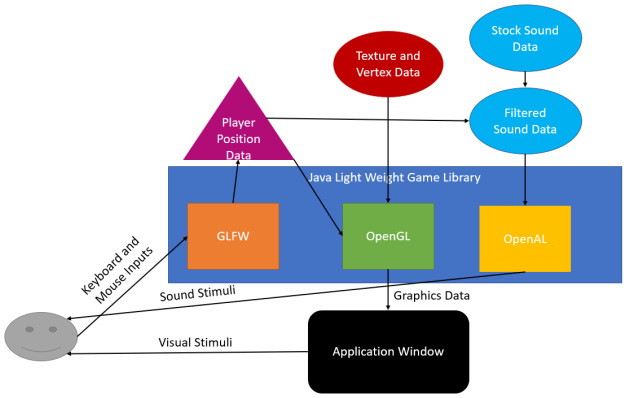

Figure 1: System Diagram

2.2.2 System. The user inputs are passed to GLFW (Graphics Library Framework) to then be processed and converted into user position and direction data within the scene. The inputs will be keystrokes of arrows for movement and two other keys for rotation. GLFW is a library that allows programmers to manage windows and OpenGL contexts and handle keyboard and mouse input [3]. The position data is sent to OpenGL (Open Graphics Library) to render the visual scene. It is also sent to be combined with the stock sound data, which includes both the sound and static position data of the sound sources (the animals). The filter data travels to OpenAL (Open Audio Library) to be processed and sent to the audio drivers, which are the interface for the user's audio devices. The user will look at the application window to view the visual stimuli. See Figure 1.

The 2D visual scene will rendered in a top-down 3rd person perspective and will use free pixel art downloaded from the internet. The art will be used for the pig, cow, rooster, monkey, bird, frog, seagull, crab, trees, barn, sand, jungle floor, grass, and ocean. The user will be able to move around the trail in 2D. Boundary collision detection will be implemented to prevent the user from exiting the trail. We will use OpenGL libraries and functions to implement this functionality The scene will be similar to Figure 2. Creating a simple visual interface will allow us to prioritize effective 3D audio rendering.

To implement the audio component, recordings of crab movement and typical noises that pigs, cows, monkeys, frogs, crabs, birds, and seagulls make will be downloaded from the internet as MP3 files, and then will be converted to PCM (Pulse Coded Modulation) format. Within OpenAL, these animals' locations in the scene will be set as sources, or locations where sounds are emitted, in the 3D audio scene [5]. The user's location in the scene is passed in to OpenAL and set as the listener, continuously. As the user moves around the visual scene using the arrow keys, the listener will move accordingly in the audio scene in OpenAL. The sources will be programmed to emit the corresponding sounds with random frequency. The user will perceive the 3D animal sounds to be coming from their relative location and direction in the rendered visual scene. The third-person person perspective may negatively affect the user's perception of 3D audio. Different camera angles and distances will be compared to determine which setting results in the most immersive audiovisual experience for the user.

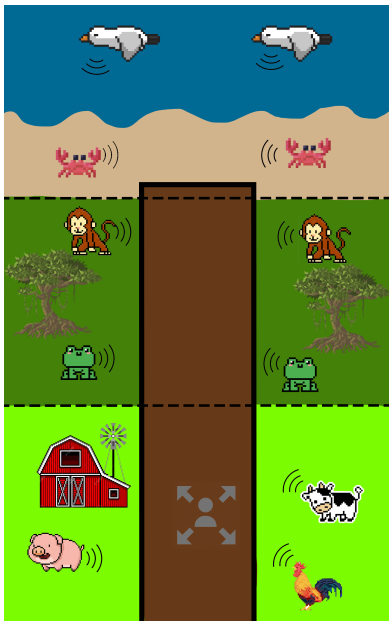

Figure 2: Visual Interface Wireframe

2.3 Procedure

The testing will take place in a quiet indoor area with the participant sitting in front of a computer. The participant should have a pair of over-the-ear headphones placed on their head with the volume at a comfortable intensity. The application will be started, and a brief

overview will be presented by the testing administrator. The user will then be allowed to navigate through the map with minimal guidance. The testing administrator will record user pauses.

After this preliminary assessment, the application should be reset, and this time the user should be instructed to go through the map and point out the animal that sounds the closest and the furthest for each side of the trail.

After the last step, the application should be shut down and the configuration file opened. The configuration file will have the type of animal and a gain value as key/value pairs. There will be no values for gain in the configuration file at first. Once the gain value for each entry is set to 0.5, the application should be launched. The application will have red squares directly in front of where the animals are that match the configuration file, and other animals will not make sounds.

For each square, the user will listen and mark on a printout map where they perceive the sound is coming from. Next, the application should be shut down, and the configuration modified, and the previous steps followed for gain values of 0.1, 1.0, and 2.0 [7].

2.3.1 Statistical Treatment. The experimental procedure will yield a few pieces of data geared toward the range and attenuation of sounds [7]. The experiment's goal is to verify that the sound filtering creates an immersive sensation to the user. The user should be able to feel spatially connected to the character and sense the environment around them.

The first set of data is the observations taken by the testing administrator when the user navigates the map with no guidance. The administrator will take notes on long pauses, facial expressions, and any backtracking to listen to specific sounds again. This gives a starting idea on which parts are most interesting or confusing. The locations of these pauses between the users will be graphed and analyzed to see there are any correlations. This will help us investigate certain parts of the map that are problematic.

The next data is the perceived range of each animal sound on the trail. The goal is to program an obvious choice, which will be verified by the testing subject. The percentage of how many users are able to select the correct animals based on the perceived spatial location of the sound of the animals will be calculated. Success will be achieved if the majority of users are able to select the intended animals.

In the final set of metrics, the condition of the application will be changed by modifying the gain of the animal sound calculations. The changes in attenuation and the effect on the perceived distance of the animals will be measured [7]. The idea is to adjust the gain to modify the attenuation of the sound from specific animals [7]. The distance on the printed-out maps from where the users marked will be measured and compared with the expected perceived distance. Success is measured by how close the markings are to the expected markings. If most of the markings from the users are less than 5 millimeters from the expected position, that will be successful.

2.4 User Survey Procedures

We will ask each of the 5 participants in the user survey the following five questions:

(1) What is your favorite computer game? Why? Can you describe the game?
(2) What animal do you think makes the coolest sound? What about the sound do you like?
(3) What animals would you expect to find on a farm? In the jungle? At the beach?
(4) Do you prefer games with background music or no background music? Why?
(5) Do you prefer games that teach you something or games that are just for fun?

2.5 User Study Procedures

Each of the five participants in the user study will run the same experiment which assesses three metrics in our application: perception of environment, localization accuracy, and usability. The assessments were performed using three test pilot routines:

2.5.1 Perception of Environment.

(1) Place the character in the center of the farm path.
(2) Wait 10 seconds.
(3) Ask the user how many types of animals they hear and what types they hear.
(4) Ask if the environment sounds like a farm.
(5) Place character in the center of the jungle path.
(6) Repeat steps 3 and 4, but instead for the jungle.
(7) Place the character in the center of the beach path.
(8) Repeat steps 3 and 4, but instead for the beach.

2.5.2 Localization Accuracy.

(1) Turn off all of the animal sounds except the chickens, the jungle bird, and the seagull.
(2) Place the character about 5 feet below the chicken, but still in the center of the path.
(3) When the chicken sound signals, have the user point to the direction of the chicken.
(4) Take note of the angle of the hand as it relates to azimuth angle and elevation as the head of the user as the reference point.
(5) Place character about 5 feet above the chicken, but still in the center of the path.
(6) Repeat step 4 and 5.
(7) Repeat step 2 through 5, but for the jungle bird.
(8) Place the character in front of the seagull.
(9) Repeat step 4 and 5 for the seagull.

2.5.3 Usability.

(1) Ask the user to walk freely and find their favorite sounding animal, take note of any difficulties.
(2) Ask the user to navigate to hear the seagull, take note of the time it takes to do that.
(3) Ask the user then to navigate back to the cow, take note of any difficulties.
(4) Ask the user to go listen to the crab, take note of any difficulties.

3 RESULTS

3.1 User Survey Results

3.1.1 Response 1.

(1) Roblox because of all the mini games in it.
(2) Cat because they make lots of different sounds that are all cute, but some are annoying.
(3) Farm: Cow, cat, pig, dog, horse, sheep, goat, chicken
Jungle: Jaguar, tiger, monkey, snake, spider, birds, frogs, tadpoles
Beach: Fish, crabs, clam, pelican, seagull
(4) It depends on how she feels but she usually likes no background music better.
(5) Just for fun, but her mom likes it when she plays games that are educational.

3.1.2 Response 2.

(1) Minecraft because there's no end to the game. A lot of people play it, so she can watch them play. It has endless possibilities and you can do whatever you want.
(2) Frogs make a pretty cool sound because it is relaxing and sort of means you're in a peaceful place.
(3) Farm: Goats, sheep dogs, sheep, cow, pig, cat, chickens.
Jungle: Pumas, leopards, rhino, monkeys, bears.
Beach: Crabs, jellyfish, clams, shrimp, lobsters.
(4) Depends on the game. If it's more realistic, it is better with no background music. Less realistic games are better with background music.
(5) Prefers games just for fun, especially story games.

3.1.3 Response 3.

(1) The Isle. It's the most realistic dinosaur game where you can actually be a dinosaur. The game is about being a dinosaur. You choose which dinosaur you are, and you can join other dinosaurs, lay eggs, and battle other dinosaurs.
(2) Favorite animal sound: dilophosaurus. Other sound: dryosaurus. Because he likes dinosaurs.
(3) Farm: cow, chicken, pig, goat, dog, sheep, mice and rats, a cat.
Jungle: boa constrictor, jaguar, monkees, tropical birds.
Beach: hermit crab, jellyfish, sea cockroaches, seagulls, fish, beetles, shrimp.
(4) Prefers cheerful music backgrounds and doesn't like creepy music. Music is fun and exciting.
(5) Prefers learning if it's about strategy.

3.1.4 Response 4.

(1) Fortnite because it's fun. You get to find things in houses and trees and campsites. Likes fighting other people and getting friends to play with.
(2) Coolest animal sound: dolphins. Also likes creek sounds.
(3) Farm: dogs, chickens, cows, pigs.
Jungle: orangutans, sloth, monkees, birds, frogs, lizards, snakes, boa constrictor.
Beach: sharks, crabs, dolphins, fish.
(4) Prefers background music because loves music and it sounds good. If it's a dramatic part, and you are doing dramatic things, it's cool.
(5) Prefer games that are just for fun.

3.1.5 Response 5.

(1) Drawing with the Paint application because that is the only computer application used yet.
(2) Favorite sound is from a horse because "it is very cool".
(3) Farm: Duck, Bird, Pig, Cow.
Jungle: Spider, Frog, Turtle, Bugs, Unicorn.
Beach: Fish, Whale, Jellyfish, Shells.
(4) Never played a game with background music.
(5) Games that are just fun.

3.2 User Study Results

3.2.1 Participant 1. Perception of Environment: The user correctly was able to identify that there were three distinct animal sounds in each of the environments, and that the sounds seemed to match the intended environments.

Localization Accuracy: The user pointed correctly towards the chicken in both cases. This should be roughly -45 degrees azimuth and elevation of 0 degrees when the chicken is in front, and -45 degrees azimuth and an elevation of approximately 230 degrees when chicken is behind. The user was slightly less accurate with the jungle bird as the azimuth is expected to be around 45 degrees, but the user said it was closer to 80 degrees. The same result was recorded when standing in front of the jungle bird. This could mean the jungle bird needs to have it's coordinates checked and verify the angles that are calculated for the HRTF calculations. The user clearly pointed directly in front of them when placed in front of the seagull.

Usability: The user had no difficulties using the arrow keys to navigate directly to the correct animals.

3.2.2 Participant 2. Perception of Environment: The user was able to accurately identify the animal sounds and the environment they were in for all 3 environments.

Localization Accuracy: The user accurately pointed toward the chicken and seagull. The user was close, but slightly more off with the localization of the jungle bird.

Usability: The user navigated to the seagull and crab very quickly, but had a slight difficulty when going to the cow. Briefly hit the corner where the two paths meet, but quickly corrected and navigated to the cow.

3.2.3 Participant 3. Perception of Environment: The user was able to recognize the correct number of animals in each zone by moving around a bit, and also the correct type of environments.

Localization Accuracy: The user was able to accurately point out the direction of the chicken in both use cases. The user also did not struggle with pointing out the jungle bird accurately. The seagull also was seen to be pointing ahead.

Usability: The user did not have any issue navigating and waiting for the animal responses around the map, although, it was mentioned that the character not having animation as it moves is a bit odd.

3.2.4 Participant 4. Perception of Environment: Was able to recognize the sounds, but was confused about the visuals; the participant was confused at first by looking at the crab, but hearing the pig.

They then realized where they were located on the map. They then were able to correctly answer which locations they were.

Localization Accuracy: The user identified the correct direction for the chicken when the chicken was in front of the character, but the complete wrong direction when behind. The direction was correct for the jungle bird and the seagull.

Usability: The participant asked multiple times to go to the jungle to hear the snake again. They however thought it was a worm at first. The user laughed when navigating with the arrow keys to the monkey. They struggled with the arrow keys as they are not familiar with how they work, and have never used them before. The participant spent a long time navigating around the map and giggling at certain animals.

3.2.5 Participant 5. Perception of Environment: The user was able to accurately identify the three animal sounds in each zone. They also identified the environment they were in for all 3 environments.

Localization Accuracy: The user accurately pointed toward the jungle bird and seagull, with a difference of less than five degrees between the angle of the hand as it relates to azimuth angle and elevation as the head of the user as the reference point. The difference was larger, about 15 degrees, for the chicken when the user was in front of the seagull.

Usability: The user had no issues using the arrow keys to navigate around each environment.

4 EXISTING COMPANIES AND SYSTEMS

4.1 Companies of Interest

Video game developers like Insomniac Games, Polyphony Digital, and others have been developing games that render 3D sound for the Sony Playstation 5 (PS5), using its Tempest 3D Audio Engine. [13] Games are being developed for the Microsoft Xbox Series X as well, which use its spatial sound features: Dolby Atmos and DTS:X (Dolby Theater Systems). [6] While these next generation consoles have the capability for producing 3D audio, it is the responsibility of game developers to develop games that produce 3D sound. Open challenges for producing high quality 3D game audio include HRTF blending when the sound source is not at a sampled location, handling moving sound sources and removing consequent phase artefacts, making sounds emit from an area rather than a point, and HRTF selection and synthesis. [12] We overcame one of these challenges in our project, so these video game development companies would likely be interested in learning how our application produces 3D audio and how the technology could be integrated into their game engines.

Game development companies such as WildWorks studios, which develop educational computer games for all ages (like Animal Jam Classic), would be interested for the same reason. [1] Virtual reality experience companies like Unity Software and Oculus would also be interested because they would benefit from being able to easily add 3D audio to any virtual reality experience they create.

4.2 Addition to Existing Systems

The 3D Audio API could be adapted to support existing 2D top-down pixel art computer games. For example, the game Animal Jam Classic, developed by WildWorks studios, is a 2D online virtual world in which users learn facts about zoology. [1] If the user's character's position and game item locations are passed to the API, then 3D sound could be rendered through headphones as the character moves around each scene.

In the same way, the 3D Audio API could be integrated into a silent disco system. Certain objects in a real-world 3D environment could play 3D audio sounds relative to each user's location as the users dance and move around.

5 FUTURE RESEARCH

5.1 An Extra Semester

This semester's project focuses on creating clear, functional 3D audio. With an extra semester, we would want to focus on improving the graphical user interface that creates the foundation for the audio experience. Firstly, we would migrate the two-dimensional environment into a three-dimensional one to improve the immersion of the experience. This would take much more time than the current 2D top-down interface we created did but would substantially improve the overall experience. With the new 3D world, we would also be able to implement character rotation so that the player can turn to face each animal from any direction instead of just directional movement.

Another major addition that we believe we could accomplish with an extra semester is animal animation. Currently, everything on the page remains static except for the player who can traverse the map. We would like to animate the animals as well to make the scene come to life. This could include small movements like a bird flapping its wings or a cow chewing grass, but could also be expanded to the animals moving freely throughout their section on the map.

To improve the audio component, we would add background music that would be themed to each environment. This would also increase the level of immersion experienced by the user and should overall make the game more lively and fun. More environments with more animals and sounds could also be added in addition to the three we currently have. To account for the extra space, each environment would be the size of one screen, and when the player exits the side of the screen, they will enter the next one.

5.2 The Sky is the Limit

Building off of the additions we would make with an extra semester, we would create many more environments with many more animals, and they would all be navigable through one central map. An educational component could be added to help familiarize the young audience with new animals in an engaging way. This could include a label popping up over each animal's head once the user gets close enough.

The user could also enter an "objective mode" in addition to our current "explore mode." In "objective mode" the user would be given a set of goals to complete throughout the map. For example, one objective could be "find a certain animal." This could also be supplemented by user actions like jump, climb, and pickup.

The scene itself would also undergo several improvements. Micro-animations throughout the application would help incorporate movement into every aspect. Small movements like the wind blowing into the grass and trees and the waves moving in the scenes

with water would have a dramatic impact on the quality of the scenery. This would also greatly affect the quality of the audio because sound would be a component of everything — not just the animals. Those slight movements would create sound, such as the sound of the wind or the waves.

6 CHALLENGES

Earlier in the semester, we anticipated having issues with communication due to the virtual setup of the course. However, we faced this issue head-on by always being active in Discord discussions and participating in regular Zoom meetings. While this was a substantial barrier to us assisting each other when we got stuck on our parts, we became very familiar with screen sharing and various doodling applications to help illustrate our issues for each other. We also knew that learning how to create 3D audio would be the biggest challenge we would face in our project since none of us had experience with it, and that was the major component of the project. This of course continued to be our biggest challenge throughout the semester as we navigated new technologies.

One challenge we faced that we had not initially anticipated was time and team management. Since we knew going into the project that we all had similar time schedules and working habits, we did not predict that this would be an issue. However, when faced with other supplemental assignments in addition to the main project, it became difficult to properly delegate everything in a way that was fair to each of us. As the semester went on, we found better ways to deal with that issue and keep up-to-date on assignments while still making progress on our project. We also had some issues with packaging with Github where the project would not compile each time someone pulled new code. This issue was frustrating because it stalled collaboration, but we were eventually able to correct the packaging. While we anticipated 3D audio would be a substantial challenge, we were not entirely prepared for the significant blocks in our progress we encountered. Getting our 3D audio code to work took many weeks and meetings with the professor, but were ultimately able to get it working.

REFERENCES

[1] [n.d.]. https://classic.animaljam.com/en
[2] [n.d.]. Central Park Zoo. https://centralparkzoo.com
[3] [n.d.]. An OpenGL library. https://www.glfw.org/
[4] 2014. 3D Animal Sounds. https://www.amazon.com/Best3DApps-3D-ANIMAL-SOUNDS/dp/B00K9YMC9O
[5] Antonio Hernández Bejarano. [n.d.]. Audio. https://lwjglgamedev.gitbooks.io/3d-game-development-with-lwjgl/content/chapter22/chapter22.html
[6] Becca Caddy. 2021. Xbox Series X Dolby Vision and Atmos: what's supported, and what isn't? https://www.techradar.com/news/xbox-series-x-dolby-vision-atmos
[7] Deckhead. 2020. Complete guide to Openal with C++ Part 3. https://indiegamedev.net/2020/04/12/the-complete-guide-to-openal-with-c-part-3-positioning-sounds/
[8] Michele Geronazzo, Amalie Rosenkvist, David Sebastian Eriksen, Camilla Kirstine Markmann-Hansen, Jeppe Køhlert, Miicha Valimaa, Mikkel Brogaard Vittrup, and Stefania Serafin. 2019. Creating an audio story with interactive binaural rendering in virtual reality. https://www.hindawi.com/journals/wcmc/2019/1463204/
[9] Veronica Kennedy. 2019. Sound sensitivity in children. https://www.entandaudiologynews.com/features/audiology-features/post/sound-sensitivity-in-children
[10] Chris McGowan. 2018. AVATAR: FLIGHT OF PASSAGE: A Cinematic, Multi-Sensory 3D Experience that Soars. https://www.vfxvoice.com/avatar-flight-of-passage-a-cinematic-multi-sensory-3d-experience-that-soars/
[11] Kyle Melnick. 2017. Get Lost In The Rainforest With 'Amazon Odyssey' VR. https://vrscout.com/news/rainforest-amazon-odyssey-vr-experience/#
[12] Roland Moore-Colyer. 2020. PS5 and 3D audio: Everything you need to know. https://www.tomsguide.com/features/ps5-and-3d-audio-everything-you-need-to-know
[13] Isabelle Tomatis. 2020. Experience PS5's Tempest 3D AudioTech with compatible headsets at launch, TV Virtual Surround Sound coming after launch. https://blog.playstation.com/2020/10/06/experience-ps5s-tempest-3d-audiotech-with-compatible-headsets-at-launch-tv-virtual-surround-sound-coming-after-launch/