Navigate as a Bat. Real-Time Echolocation System in Virtual Reality

Andreasen, Anastassia; Zovnercuka, Jelizaveta; Konovalovs, Kristians; Geronazzo, Michele; Paisa, Razvan; Serafin, Stefania

Published in:
Proceedings of the 15th International Sound & Music Computing Conference (SMC2018)

DOI (link to publication from Publisher):
10.5281/zenodo.1422555

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

NAVIGATE AS A BAT. REAL-TIME ECHOLOCATION SYSTEM IN VIRTUAL REALITY

Anastassia Andreasen, Jelizaveta Zovnercuka, Kristian Konovalovs, Michele Geronazzo, Razvan Paisa, Stefania Serafin
Dept. of Architecture, Design, and Media Technology
Aalborg University Copenhagen
{asta,mge,rpa,sts}@create.aau.dk, {jzovne13,kkonovalv13}@student.aau.dk

ABSTRACT

Several studies provide evidence that blind people orient themselves using echolocation, transmitting signals with mouth clicks. Our previous study within embodiment in Virtual Reality (VR) showed the possibility to enhance a Virtual Body Ownership (VBO) illusion over a body morphologically different from human in the presence of agency. In this paper, we explore real-time audio navigation with echolocation in Virtual Environment (VE) in order to create a feeling of being a virtual bat. This includes imitation of the sonar system, which might help to achieve a stronger VBO illusion in the future, as well as build an echolocation training simulator. Two pilot tests were conducted using a within-subject study design, exploring time and traveled distance during spatial orientation in VE. Both studies, involved four conditions – early reflections, reverb, early reflections-reverb (with deprived visual cues) and finally vision. This resulted in preferred reflection pulses for the test subjects with musical background, while only reverberation features were favored by non-musicians, when being exposed to VE walking-based task.

1. INTRODUCTION

Virtual Reality (VR) has the potential to provide test subjects with a compelling experience of flying [1, 2], while inhabiting the body of a flying creature. Bats are the only existing flying mammals in nature that have anatomical similarities to humans body structure [3], which gave us a possibility to study the influence of morphological different body on human perception of Virtual Body Ownership (VBO) illusion in virtual environment (VE) [4] as well as possessing an ability to echolocate [5]. In our ongoing research, we are exploring the amount of VBO illusion that test subjects experience over a virtual bat's body through visuotactile stimulation of their arm while seeing an object touching their virtual wing, and exploring agency, testing voluntary limbs movement of the bat through VE. A linear relationship between agency and VBO was noticed. The higher agency was the more VBO was experienced by test subjects. The novel aspect of this research is an attempt to enhance VBO illusion with the help active voluntary movements during echolocation applying user-generated mouth clicks. We applied VR technology to navigate with virtual avatar, presented in Figure 1.

Bats have both day- and night-time visions, and it is considered that they see in black and white gamma, however details of objects are missing in their visual system [5]. In order to be able to get the details of the objects and to orientate themselves in the dark bats are using echolocation scope. Echolocation is ability to navigate in space through the analysis of the emitted sound pulses (echo), such as reflections and reverberations [6]. Though bats are complimenting both of their abilities, such as using sonar while capturing prey, and vision when spatial navigating and orientation in the larger environments [7]. When they have to capture a prey they simply generate the sound and their cochlea system analyses emitted impulses of the sound waves. The sound itself is generated in larynx of the bat, though Rousettus (mega bat) does it by clicking the tongue. Their pinna acts as a horn. The larger pinna they have the better low frequencies have been transmitted [5]. Through echolocation bats decode: target direction (echo frequency), distance (pulse-echo time delay), angular direction (ear amplitude difference), velocity and trajectory (pulse-echo frequency change), target size, shape as well as materials (echo frequency) [5, 6].

When offering people an experience of "seeing" the world through bat's eyes, it might assist them in learning audio spatial navigation being in a safe and controlled VE sys-

Figure 1: Visualization of the bat used in the simulation.

tem, as well as give an opportunity to perceive surrounding environment as bats' do, which might get us closer to answering such a philosophical question as "What is it like to be a bat" [8]. The described system is a part of ongoing research seeking to address the following challenges related to this illusion: (1) providing users with the sense of ownership of the bat's body; (2) enabling users to control the bat's body in a way that is both intuitive and mimics real bat's behavior; (3) offering users with impression of how the world "looks" when experienced through bat's senses; (4) affording users with the sense of control over their active movements and intentions (agency) during their flight in VE using echolocation for spatial navigation. For this purpose, we imitate bat's basic physiological ability to flight – by inhabiting its body, and orient in the environment – by implementing physical-based flight algorithm together with echolocation system. Research results revealed a larger difference in spatial navigation abilities between the participants based on their musical background: musicians and non-musicians. Musicians were able to differentiate sound pulses from reflections, while non-musicians were not able to do so. Non-musicians oriented themselves in VE mostly based on reverberations. This is a new insight on the problem of learning and training echolocation abilities among humans, which this study will address and analyze in details in the results and discussion sections.

2. ECHOLOCATION BACKGROUND

Almost 230 year ago Lazzaro Spallanzani started studying echolocation, but only approximately 78 years ago Donald Griffin offered his contribution in echolocation as we know it due to development of new at that time technology, being able to record and detect and analyze high-frequency pulses produced by bats [9]. Bats are able to perceive, distinguish and analyze different sound qualities and understand the surroundings, which is helpful in hunting and navigation. Most animals echolocate by producing different types of sound waves. For example, bats emit sonar signals in the range between 15 to 200 kHz [10]. However, some bats produce tongue-clicks to navigate in space [5], which was adopted by humans [11]. Though some studies still use whistles, hisses, speech along with white noise, or other emitting sounds from the objects [7] as well as assistive technology. However, studies suggest that echolocation ability could be reduced when using assistive devices [12, 13].

Blind people use mouth click based echolocation to help themselves navigate in unknown environments. In this regards, both spatial (emission sound pulse) and time domain (level of details) should be considered. By producing specific short clicks, they are able to perceiving direction, distance, size of the objects and shape along with material properties through pulses emitting back from the objects [7, 11, 14]. Preferred mouth click by blind people is determined to be approximately 3ms long, with a frequency component around 2-4 kHz [11]. Nonetheless, the working mouth clicks could have a range of 3-15ms with peak frequency of 3-8 kHz [7]. The sound propagation has a cardioid directivity pattern [11], while the traveling distance of the sound is 2-10 meters [15].

In order to spatially orient in the environment by detecting reflecting surfaces one should take into consideration azimuth, i.e. the time and amplitude differences in sound reaches the ear (horizontal angle), elevation (vertical angle), e.g. navigation during flight in case of bat simulations, distance to objects/obstacles, and velocity, e.g. if bats are moving and the acoustic size of an object, which might have similar acoustic size regardless distance [7, 16]. Sighted people have an ability to detect difference in about 40 cm at a distance of 170 cm, 80 cm at a distance of 340 cm, 125 cm at a distance of 680 cm away from the object [7]. When perceiving distance, it is possible to assume intensity of the direct sound over intensity of the reverb sound from surrounding surface – the further away perceived sound is the smaller the ratio should be [7].

In general echolocation is considered to be "the second vision" among blind echolocation experts, as several MRI results showed activation of visual cortex during auditory processing [7]. As they practice their echolocation abilities, they are better in retrieving information about reflections and reverberation from surrounding surfaces due to their induced higher auditory skills than among non-trained sighted people. Results show that blind people posses extraordinary ability of auditory navigation,retrieving information from the environment, as absence of visual cues enhance auditory perception [17]. Furthermore, in absence of visual cues audiomotor feedback is received from auditory spatial cues derived from head and body movements [17]. Based on the described skill to rely on their hearing, people with musical background might be the closest group that have similar developed ability to distinguish more sound variations, analogous to blind people, as they train that skill for years. Results have shown that sound plays a key role on emotional judgment among musicians, when they have to review audio information in the presence of visual cues [18]. Thus musicians might be more accurate in distinguishing echo pulses due to their ability to better differentiate audio signals due to training and their developed perceptual ability to process audio-visual information differently from non-musicians.

3. MOTIVATIONS FOR ECHOLOCATION IN VR

Several attempts have been made to build VR echolocation systems for teaching purposes [19], for architectural space exploration [20], for simulating bat's sonar abilities [10], however virtual navigation was rendering only acoustic information excluding visual presentation of the body and body perception. Sound in VR requires taking into account several acoustic cues, such as the size of the room and its architecture through reverberation, materials through reflections from the surfaces, sound source localization if such exist in VE, etc. (see [21] for a recent review on these topics). All this might help to reconstruct a cognitive spatial map for the user and help to navigate better in VE.

The majority of the studies regarding echolocation are using categorical tasks approach of psychophysical measurements, introduced by Kellogg in 1960 [16], who also systematically studied mouth clicks long with other sounds. This approach measures the ability of the participants to identify specific points from a controlled number of introduced options based on tasks completion.

It has been noticed that is useful to have task relating voluntary movements (self-motion) of the head and the body, as involving vestibular and proprioceptive input have a stronger positive influence on perceptual accuracy of echolocation [16, 22]. These studies indicate that when voluntary movements are occurring (e.g. test subject is passing along some audio source), audio scene could be considered as streaming paradigm, partially determined by spatial cues as well as by perceptual re-organization after motion. This means that as soon as an acoustic change happens at the ears, interpretation is being reset. Moreover, it was concluded that sighted subjects are able to learn echolocation [23].

Though it is more difficult for a human being to process all the information echolocation contains, as we are able to learn it only during our single lifetime, while bats have developed echolocation abilities through their whole existence as species as a result of evolutionary process [5]. Studies revealed that people tend to suppress echolocation information in in favor to the primary signal and its direction, which is called a "Precedence Effect" – getting a direct information from the sound source, especially when visual cues are present in the scene [23]. The rest of the information, such as reflections and reverberation is conflicting with the primary signal due to the time lag, thus often being disregarded due to our physiological disability to depict everything from the environment, as bats do. Though people with musical background are less receptive to interference from visual information flow, when e.g. judging auditory emotions [18]. Nonetheless, the ability to depict all possible signals from the environment can be trained. Therefore, building a controlled VR echolocation system might be useful for developing and mastering echolocation ability safely. In order to try to avoid suppression effect it might be possible to visualize sound, taken into consideration reflection frequencies and absorption properties of the materials, as well as time delay and head-related transfer function (HRTF). HRTFs allow us to identify objects' position in space, simulating acoustic transformation introduces by the human body in sound spatialization technologies for VR [24]. Moreover, reverberation algorithms could be used as a tool to provide information about room size and materials, e.g. bigger virtual room would create bigger amount of reverberation, while the loudness of the reverberation will depend on the reflectivity and absorption characteristics of room materials [7].

As a results of these motivations, in order to create a successful echolocation application, the following features will be used in VE for spatial navigation during flight locomotion [15]:

- head-related transfer function (HRTF),
- time delay from the sonar signal emission to detection the reflections,
- amplitude (loudness) and frequency of room/object reflections,
- absorption property of materials.

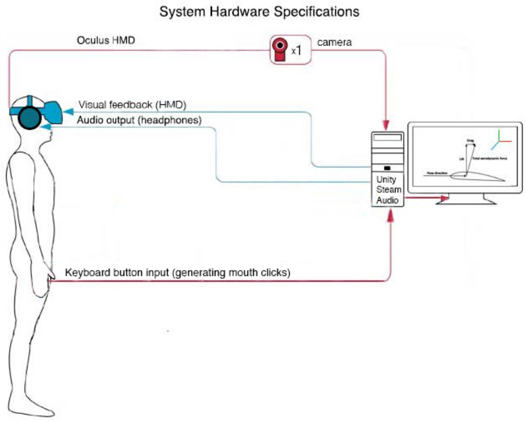


Figure 2: Schematic view of the system set-up.

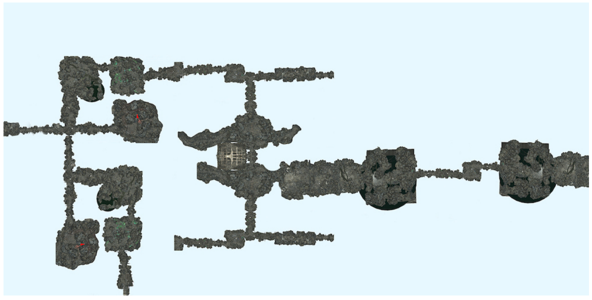


Figure 3: A top view map of the tunnel.

4. SYSTEM ARCHITECTURE

The system consists of an Oculus Rift head-mounted display (HMD). Headset resolution is 2160×1200 at refresh rate of 90Hz and FOV $110°$. One Oculus camera is used for HMD positional tracking. For the echolocation system, users were able to generate pre-recorded mouth clicks by pressing a "space" button on the keyboard. The audio stimulus was delivered through semi-closed Razer SWTOR Gaming Headset headphones. The overall system setup is presented in Fig. 2.

4.1 Sound propagation of mouth-clicks and echolocation

A virtual cave scene was developed in Unity 3D (version 2017.3.0f3). The virtual cave environment was comprised of multiple stone tunnels different in length, closed on one end and leading to the large rounded stone cave. A top view map of the tunnel is visible in Figure 3. The cave environment is presented in Figure 12.

As a real-time echolocation is essential for consistent and believable experience of VE, especially when introducing morphologically different virtual body, the whole setup allowed real-time calculations, providing information about the size of the objects and VE materials, depending user

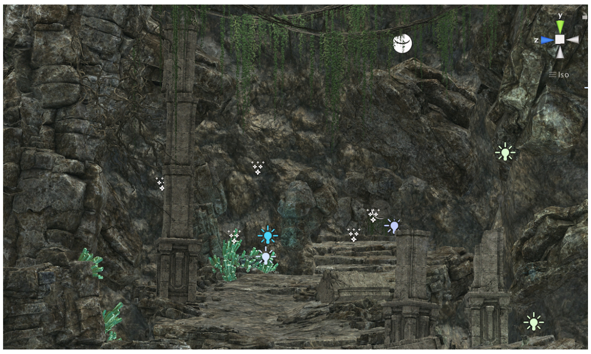


Figure 4: The virtual cave.

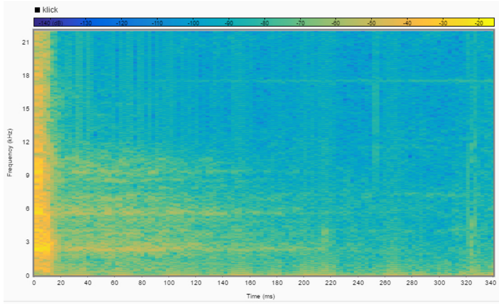


Figure 5: Spectrogram of the pre-recorded mouth clicks.

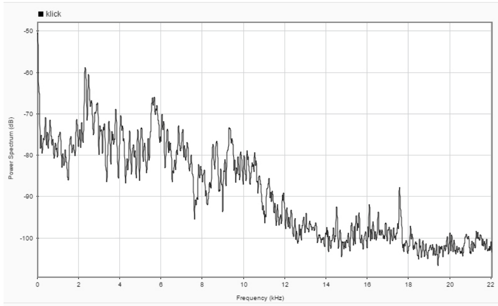


Figure 6: Visualizations of peaks in the frequency response of a pre-recorded click.

position in space. Steam Audio plug-in for Unity 3D allows to implement spatial audio design applied in VR and therefore was used to simulate natural sound propagation by calculating reflections and occlusions of the input mouth clicks. The following settings were used in the scene:

– Bilinear HRTF interpolation;
– Physics-based attenuation;
– Air absorption;
– Direct mix level = 1.0 (minimum was 0, maximum was 1);
– Frequency dependent transmission occlusions using partial method, e.g. casting multiple rays from the sound source within the radius of 5 meters;
– Real-time reflections with indirect mix level = 0.7; (minimum = 0, maximum = 16. This variable presents indirect sound reflections with HRTF-based 3D audio).

Unfortunately, it was impossible to switch off some of the reflection levels, as Steam Audio did not allow it without switching off all the reflections at once. Additionally, real-time late reverberation was added through the mixer. Every object had material-based audio shader with specific parameters of sound absorption and reflection. Material presets used in the scene were the following: "Rock" for the walls, "Wood" for the trees, roots and vegetation, "Glass" for crystals. Space button has been chosen as a trigger for playing pre-recorded mouth click samples instead of user generated clicks though microphone which introduces a latency of 200ms once processed by Unity. Mouth click was recorded based on the features discussed in Sec.2. The length of pre-recorded mouth click was 13ms and frequency component was the highest at 2-3 kHz with peaks on 6 kHz and 9 kHz.

Pre-recorded mouth clicks, based on the above-mentioned feature could be seen in Figure 5, which characterizes length of the click, frequency and loudness (see also Fig.6 for the frequency response).

This setup allowed us to test the Steam Audio engine in an immersive experience into VR, where participant would be able to spatially navigate and orient themselves aided by the implemented audio features.

5. METHODS

The main goal of this preliminary assessment was to find out if participants were able to orient themselves in a VE with the help of audio cues and to exploit sound qualities that might influence positively the navigation experience. On the other hand, the technical goal was to evaluate the information provided by Steam Audio, mainly if participants might be able to perceive the auditory qualities of an environment with this engine, and understand chosen spatial audio cues, characterizing an echolocation task.

Accordingly, We tested four conditions for VR navigation: (i) unimodal visual feedback, (ii) unimodal auditory feedback based on reflection rendering only, (iii) unimodal auditory feedback based on late reverberation only, and (iv) unimodal auditory feedback based reflections plus reverberation without vision. For this purpose, within subject study design ($N = 8$) was conducted, involving 7 males and 1 female, ages between 22-30 ($M = 26, SD = 2.39$). Participants were sampled randomly. Tests were conducted at the Multisensory Experience Lab, Aalborg University Copenhagen in noise-controlled room. The majority of the test subjects were not musicians (5/8), had normal hearing and no spatial orientation disabilities. Three of the test subjects were trained as semi-professional musicians.

Conditions	Reflections	Reverberation	Vision
Reflection (RF)	*x*	*absent*	*absent*
Reverberation (RV)	*absent*	*x*	*absent*
Reflection-Reverberation (RR)	*x*	*x*	*absent*
Vision (V)	*absent*	*absent*	*x*

Figure 7: The experimental conditions.

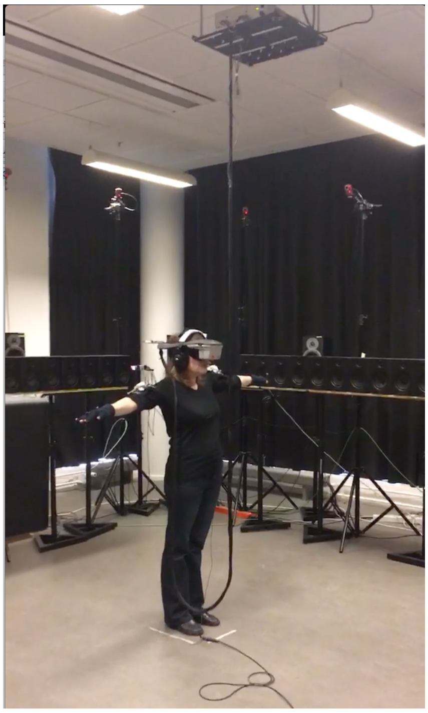


Figure 8: A testing session.

Participants were given a training session of the system with the purpose of getting acquainted with the changes offered by sound feedback, depending on the direction and distance to the objects, presented in Fig. 8.

The training session lasted up to 7 minutes. Subjects were able to experience echolocation by triggering mouth clicks and listening to the audio feedback from the virtual space through Razer SWTOR Gaming Headset headphones. The duration and amplitude of the feedback allow participants to determine primarily the distance to the objects in virtual space and, after a training period, to distinguish additional information on the shape and size of the objects in the virtual cave.

After that, participants started the experiment which was task based: they were asked to find their way out of the corridor to a bigger hall of the virtual cave. They were placed at the same position in each of the four test conditions inside the virtual cave corridor with random rotation for each condition in order to avoid learning effect in orientation. +

In order to assess participant ability to orient themselves in the environment feedback conditions were testing time passed in seconds from the beginning of the test till the reached task goal, and distance walked in meters from the beginning of the test till the end goal.

6. RESULTS

Inferential statistics was performed to find out if the difference between conditions was statistically significant. Two tests were conducted using one-way analysis of variance (ANOVA). One test measured the amount of time passed in seconds from the beginning of the trial till the goal-reached event. Another test showed the amount of distance in meters, test subjects went, to find the way out of the virtual corridor.

6.1 Time-based experiment (I)

The first test showed that the difference in reported number of seconds was statistically significant between RF, RV, RR, V, [F (3, 28) = 4.35, $p < 0.05$]. Post hoc comparison using the Turkey HSD test with confidence level p¡0.05 revealed that the mean score for condition RF ($M = 121.88, SD = 80.56$) was significantly different from condition V ($M = 16.25, SD = 3.96$), and the mean score for condition RR ($M = 144.63, SD = 112.39$) was significantly different from condition V ($M = 16.25, SD = 3.96$). Results of the first test are presented in Figure 9a.

6.2 Distance-based experiment (II)

The second test showed no significant difference in reported number of meters at the $p < 0.05$ level between $RF, RV, RR, V, [F(3, 28) = 2.66, p < 0.068]$. This could have happened due to participants ability to stay on one place without moving while listening to mouth clicks, therefore moving distance might be considered insignificant. Results of the second test are presented in Fig. 9b.

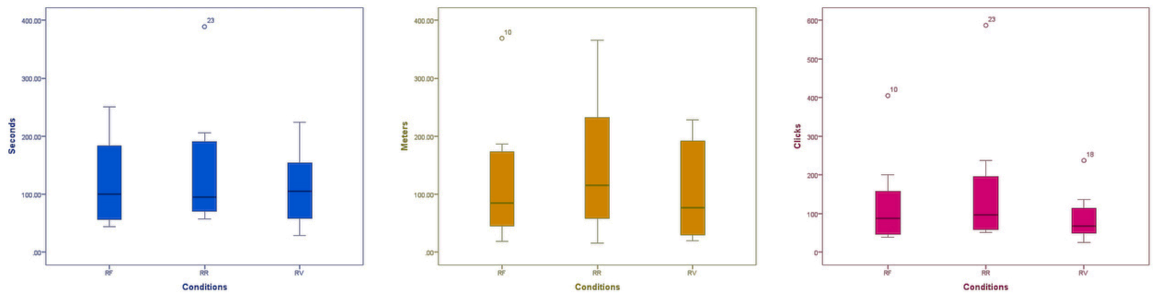
In both tests the vision condition took less ($M = 16, 25$, $SD = 3.96$) as well as less distance ($M = 23, 88, SD = 2.59$) to finish the path to the opened virtual space in the cave.

6.3 Number of clicks

Results of the analyzed number of clicks are presented in Figure 9c. The less number of clicks were presented in the RV condition, while in RR test subjects made the most number of clicks. According to participants' comments they had to make more clicks in RR condition due to difficulty in orientation, as reflections were mudded by reverberation; on the contrary, reverberations were extinguished by the low frequencies reflections thus creating mixed effects.RV condition required less number of clicks, meaning that for the majority of participants, reverberation-only condition allowed a smoother navigation in the environment.

Furthermore, Fig. 10 shows the values of mean and standard deviation among musicians and non-musicians in all four conditions. It has to be noticed that people with musical background could orient themselves in the environment without visual cues faster than non-musicians. Results suggest that musicians rely more on reflections, while non-musicians on reverberation. This might be, because reflected sound was more clear for the musicians. It allows them to distinguish certain/low frequencies that hold information about orientation and therefore allowed to distinguish distance more clear.

For non-musicians, preferable choice was reverberation condition, as it helped them to spatially orient better in the virtual environment and distinguish the size of the room, despite the location (virtual corridor or an opened space). Comparison of means for musicians and non-musicians for



(a) (b) (c)

Figure 9: (a) Results of the first experiment, (b) results of the second test, and (c) number of clicks per condition.

Time-based experiment (I) are shown in Figure 11. Musicians performed best in RF condition (mean = 61.7), while non-musicians performed best in RV condition ($M = 110.2$). The worst musicians' performance has been obtained in RV condition ($M = 112.3$), though non-musicians performed worse in RR condition ($M = 178.6$). Musicians accomplished slightly better in V condition (mean for musicians = 15.7, mean for non-musicians = 16.6). This might have happened due to learning effect, as this condition was tested at the end and was not randomized as other conditions to avoid bias.

The same distinction can be considered also for distance-based experiment (II), presented in Fig.12. All test subjects scored best results in RV condition (mean for musicians = 26, mean for non-musicians = 144.4). Musicians' mean was worse in RF condition ($M = 140.66$), while non-musicians showed worse results in RR condition ($M = 201.8$). Similar to the previous test musicians acted slightly better in V condition (mean for musicians = 24, mean for non-musicians = 24.2), which might have happened due to the above described reason.

7. GENERAL DISCUSSION

The described research had two goals in mind: (i) to analyze the perceptual ability of users to spatially navigate inside the built VR system with the help of auditory cues; (ii) to identify technical limitations of the system for navigation in a virtual cave.

7.1 Perception of audio cues

First of all, it has to be noticed that the used mouth clicks could be considered to be long (13ms with higher peak frequency of 2kHz). Blind experts mouth clicks are typically shorter (3ms), higher, more intense [11], as they are

Groups	Values	RF	RV	RR	V
Musicians	Mean	140.66	26	77.66	24
Non-Musicians	Mean	177	144.4	201.8	24.2
Musicians	SD	18	119.8	50.29	1.15
Non-Musicians	SD	116.57	76.82	123.39	3.27

Figure 10: Descriptive Statistics values for Musicians and Non-Musicians.

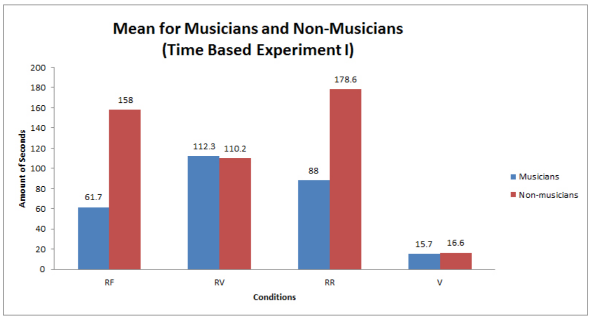


Figure 11: Comparison of means for musicians and non-musicians in the first experiment.

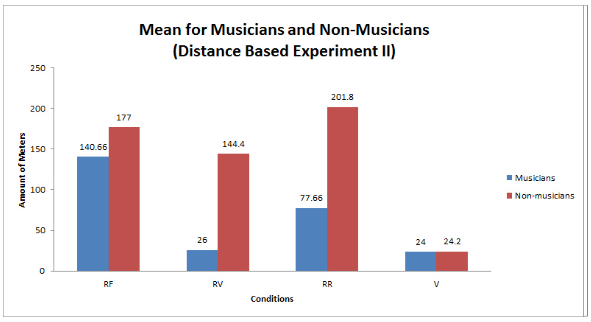


Figure 12: Comparison of means for musicians and non-musicians in the second experiment.

more experienced in perceiving and processing echo. The shorter the pulse is the more distinct echo might be perceptually for the human ear. The quality of the used click might have been harder. Therefore, the factors influencing the duration of the clicks could also be researched further. Additionally, clicks amplitude should be also investigated. And finally, the dynamic length of the click should be tested as well, e.g. bats do also use shorter and longer clicks, depending on the target. Again, some sounds, generating echo, are better for detecting the position of the objects. Nonetheless, the shape of the objects could be detected with a different sounds than the distance to that object. Thus, for orientation in general (as with the help of different frequencies it is possible to detect distance), clicks could be longer, but for detecting the objects size in front of the test subjects, shorter clicks could be used, as ration of direction to reflected sound might be more important. Therefore, for future work intensity, duration and frequency should be considered.

Reverberation time will have more attention in future experiments. RR condition resulted in the worst performance between both musicians and non-musicians, which might be due to some technical limitations of the system (see the next section for details). RV condition was preferred by non-musicians, as they mentioned that it was easier for them to depict the size of the room with only late reverberation present (while musicians preferred only reflections without reverberation). Nonetheless, it is not enough to detect only the room size, but preferably also the distance to the object, which could be detected by higher frequencies. The shorter reverberation time is (depending on reflectivity of the sound from surfaces) the more distinguishable the objects might be. Therefore, emitting properties of the materials should be investigated in depth. Accordingly, one of the fields of interest would be the difference in echolocation-based orientation depending on the quality of the materials inside VE. Since the test was performed mainly with the "rock" material applied of a virtual cave, several materials and their properties should be studied more also in terms of perceiving the differences among the objects as well as distance to them.

The last but not least consideration is related with self-motion, or in our scenario flying locomotion that might increase echolocation among users. This condition was not tested this time, because the scope of this study was on the sound qualities and technical implementation. Hence, the next step in our study is to test agency (echo-acoustic orientation, using vestibular and proprioceptive cues, as full body movement should influence on accuracy navigation and better distance detection) and its influence on VBO illusion and perception of the VE in relation to presence. As echolocation is activating parts of the brain responsible for vision, we are also implementing echolocation visualization, trying to visualize sound propagation in space, its behavior and qualities, which might help to learn echolocation.

Obtained results have shown some behavioral patterns and differences between musicians and non-musicians. Qualitative observations have led to the conclusion that musicians are more sensitive to changes in sound qualities and echoes. Two out of three musicians noted that it was better for them to navigate inside the environment detecting low frequencies. In general test subjects reported sound being "blurry" in the absence of obstacles in front of them. Differences in number of seconds between musicians and non-musicians might be due to their ability in distinguishing differences in sound qualities. Interestingly, musicians were first performing clicks in every direction, standing on the same spot and only then chose their path. Non-musicians started often moving at the same time as performing clicks. Therefore, they were actively changing their location in comparison to musicians. The changes in non-musicians position and environment around them could become confusing and the differences in reflection became harder to perceive. Changes in distance can be interpreted based on the same behavioral principles. Finally, the bigger the number of clicks was, the worse the performance of the participants became. If the number of clicks increased and the time between clicks became smaller, echoes and reflections became less distinguishable. Albeit the large number of clicks might have influenced results in a less attracted way, echo suppression among non-musicians might have also influenced their performance, as reflections and reverberations might have conflicted with the primary signal due to its natural time delay, and therefore might have been disregarded.

7.2 Technical limitations of the system

The overall system performance produced no delays, which were perceptually influent and was working smoothly with no irregularities reported by users. Unfortunately, the used software did not allow to limit the number of reflections and to control them, therefore we were working on implementing reflections with custom algorithms, e.g. using ray-casting from the audio source component of the camera directly in Unity scene.

It has to be stressed that the most important feature for our research is providing a system able to support real-time recorded mouth-clicks with a microphone input. Sadly, Unity 3D built-in features gave a delay of more than 200ms, for a generic microphone input. For the future, we will implement real-time microphone input with different technologies such as FMOD audio plug-in for Unity 3D, which is compatible with Steam Audio and could be controlled outside of Unity environment. Preliminary tests showed no delay using this plug-in. Providing users with the opportunity to produce their own mouth clicks in real-time could allow further investigation on real-time echolocation in VR and VBO.

8. CONCLUSION

The focus of this pilot experiment was to test if Steam Audio engine was able to provide auditory features in our already existing bat simulator in order to develop a perceptual experience of echolocation in VEs. Results reported a significant difference between the performances among the four conditions (reflections, reverberation, reflections together with reverberation and vision), uncovering interesting difference between musically-trained people and people without trained musical hearing. Their performances differed across the conditions with distinct navigation strategies.

Further studies will be conducted combining embodiment of the virtual bat with real-time echolocation using microphone input, taking into consideration training session time, audio quality of the clicks, materials in the environment, and the amount of the reflections (1^{st}, 2^{nd}, etc.). A statistically meaningful pool of participants will allow to compare musicians and non-musicians groups. Finally, the interaction between a visual presentation of echoes in combination with sound feedback will be considered for training users to human echolocation.

Acknowledgments

This paper was partially supported by Nordforsk's Nordic University Hub Sound and Music Computing, project number 86892

9. REFERENCES

[1] E. Sikström, A. De Götzen, and S. Serafin, "The role of sound in the sensation of ownership of a pair of virtual wings in immersive vr," in *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*. ACM, 2014, p. 24.

[2] P. Lelyveld and U. Entertainment, "Virtual reality primer with an emphasis on camera-captured vr," *SMPTE Motion Imaging Journal*, vol. 124, no. 6, pp. 78–85, 2015.

[3] J. E. Hill, *Bats, a natural history*. In the Hands of a Child, 1984, vol. 877.

[4] S. Andreasen, Niels C. Nilsson, "Agency enhances body ownership illusion of being a virtual bat," 2018.

[5] M. B. Fenton and N. B. Simmons, *Bats: a world of science and mystery*. University of Chicago Press, 2015.

[6] D. E. Purves, G. J. Augustine, D. E. Fitzpatrick, L. C. Katz *et al.*, "Neuroscience." 1997.

[7] L. Thaler and M. A. Goodale, "Echolocation in humans: an overview," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 7, no. 6, pp. 382–393, 2016.

[8] T. Nagel, "What is it like to be a bat?" *The philosophical review*, vol. 83, no. 4, pp. 435–450, 1974.

[9] M. B. Fenton, A. D. Grinnell, A. N. Popper, and R. R. Fay, *Bat bioacoustics*. Springer, 2016, vol. 54.

[10] D. Waters and H. Adulula, "The virtual bat: Echolocation in virtual reality." Georgia Institute of Technology, 2001.

[11] L. Thaler, G. M. Reich, X. Zhang, D. Wang, G. E. Smith, Z. Tao, R. S. A. B. R. Abdullah, M. Cherniakov, C. J. Baker, D. Kish *et al.*, "Mouth-clicks used by blind expert human echolocators–signal description and model based signal synthesis," *PLoS computational biology*, vol. 13, no. 8, p. e1005670, 2017.

[12] L. Thaler and J. Castillo-Serrano, "Peoples ability to detect objects using click-based echolocation: a direct comparison between mouth-clicks and clicks made by a loudspeaker," *PloS one*, vol. 11, no. 5, p. e0154868, 2016.

[13] Z. Haraszy, D.-G. Cristea, V. Tiponut, and T. Slavici, "Improved head related transfer function generation and testing for acoustic virtual reality development," in *Proceedings of the 14th WSEAS international conference on Systems: part of the 14th WSEAS CSCC multiconference*, 2010, pp. 411–416.

[14] N. Constant and L. Roberts, "Narratives as a mode of research evaluation in citizen science: understanding broader science communication impacts." *Journal of science communication.*, vol. 16, no. 4, p. A03, 2017.

[15] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature inspired cooperative strategies for optimization (NICSO 2010)*. Springer, 2010, pp. 65–74.

[16] W. N. Kellogg, "Sonar system of the blind," *Science*, vol. 137, no. 3528, pp. 399–404, 1962.

[17] J. Lewald, "Exceptional ability of blind humans to hear sound motion: implications for the emergence of auditory space," *Neuropsychologia*, vol. 51, no. 1, pp. 181–186, 2013.

[18] J. Weijkamp and M. Sadakata, "Attention to affective audio-visual information: Comparison between musicians and non-musicians," *Psychology of Music*, vol. 45, no. 2, pp. 204–215, 2017.

[19] D. ROBINSON and G. KEARNEY, "Echolocation in virtual reality," 2016.

[20] L. Picinali, A. Afonso, M. Denis, and B. F. Katz, "Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge," *International Journal of Human-Computer Studies*, vol. 72, no. 4, pp. 393–407, 2014.

[21] S. Serafin, M. Geronazzo, N. C. Nilsson, C. Erkut, and R. Nordahl, "Sonic Interactions in Virtual Reality: State of the Art, Current Challenges and Future Directions," *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 31–43, 2018.

[22] D. Hope, C. Gries, W. Zhu, W. F. Fagan, C. L. Redman, N. B. Grimm, A. L. Nelson, C. Martin, and A. Kinzig, "Socioeconomics drive urban plant diversity," *Proceedings of the national academy of sciences*, vol. 100, no. 15, pp. 8788–8792, 2003.

[23] L. Wallmeier, N. Geßele, and L. Wiegrebe, "Echolocation versus echo suppression in humans," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 280, no. 1769, p. 20131428, 2013.

[24] M. Geronazzo, "User Acoustics with Head-Related Transfer Functions," in *Encyclopedia of Computer Graphics and Games*. Springer, Cham, 2018, pp. 1–5.