VIRTUAL REALITY PLATFORM FOR SONIFICATION EVALUATION

Thimmaiah Kuppanda[1], Norberto Degara[1], David Worrall[1], Balaji Thoshkahna[1], Meinard Müller[2]

[1] Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany
[2] International Audio Laboratories, Erlangen, Germany

thimmaiah.kuppanda@iis.fraunhofer.de

ABSTRACT

In this paper we propose a game-based virtual reality platform for evaluation of sonification techniques. We study the task of localization of stationary objects in virtual reality using auditory cues. We further explore sonification techniques and compare the performance in this task using the proposed platform. The virtual reality environment is developed using Unity3D (game engine) and an Oculus Rift, a head mounted virtual reality display. Parameter mapping sonification techniques are employed to map the position of the object in virtual space to sound. Hence, the framework defined here constitutes an auditory virtual reality environment. This auditory display interface is subjectively evaluated in stationary object localization task. A statistical analysis of the subjective and objective measures of the listening test is performed resulting in a robust and scientific evaluation of the sonification methods.

1. INTRODUCTION

Sonification is an increasingly common approach in typical tasks like source localization, obstacle avoidance and navigation, hence its significance in the field of auditory display research. Sonification methods have potential application to navigation systems in vehicles and smartphones, assistive technology for the visually impaired and other eyes-free applications. The aim of these technologies is to deliver location-based information to support navigation through sound. This is a very challenging task. The main challenge is to design a meaningful auditory display that is able to communicate relevant aspects of complex visual scenes, where psychoacoustics and aesthetics are important design factors [1]. The resulting sound must be accurate in terms of the location-based information communicated, intuitive and as well as be acoustically attractive to the user. A number of different sonification methods for assisted navigation can be found in the literature [1]. In general, these methods scan the space for obstacles and synthesize the properties of the scene using various sound rendering techniques [2, 3, 4].

Hermann's definition of sonification [5] implies that sonification is a data-dependent generation of sound using a systematic, objective and reproducible transform. According to this definition, sonification can be considered as a well-defined scientific method. Both subjective and objective evaluation are important steps in the design and implementation of auditory displays and the encompassing sonification technique [6]. Nevertheless, a robust evaluation and scientific comparison of sonification methods is often neglected by auditory display researchers [7]. To address these limitations, we present a game-based virtual reality (VR) framework for a formal comparison of sonification methods for target localization. VR is a computer-based technology that provides visual, aural and tactile stimuli of a virtual world generated in real time [8]. VR has developed from research to a tool for both entertainment and training. VR is a part of wearable technology that made a major break through with availability of Oculus Rift—a head mounted display device—for VR gaming.

This paper reports on research aimed at demonstrating the use of a VR platform for the evaluation of certain simple sonification techniques. We focused on the task of localizing stationary objects in a VR environment using auditory cues. We further explored sonification techniques and compared the performance in the localization task using the platform.

The remainder of the paper is organized as follows: Section 2 describes the task that formed the basis for the proposed model. We describe the proposed model in Section 3. Section 4 explains the experimental setup facilitated by the proposed model. Section 5 presents the results of the evaluation. Finally, Section 6 draws some conclusions and considers the possible future work.

2. TASK DESCRIPTION

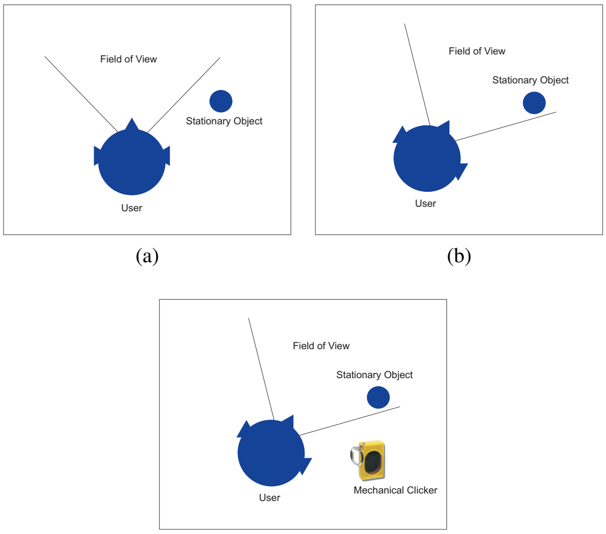
Our aim was to evaluate the performance of different sonification methods in the task of object localization. The test subject was required to find a stationary object placed in a virtual space using an auditory cue. Figure 1 depicts this task. We placed the subject at a fixed position in virtual space and the subject was able to turn 360 degrees at this fixed position. The sound conveyed information about the position of the object. Precisely, the subject was required to bring the object within the (FOV). Once localized, the subject was required to respond using a mechanical clicker, which was recorded as an objective measure of response time.

3. PROPOSED MODEL

This section describes the game-based VR framework. Section 3.1 gives an overview of the model, followed by the details of visual and auditory components in Section 3.2 and Section 3.3 respectively.

3.1. System Description

Figure 2 shows the block diagram of the proposed model. The model is comprised of the following components:



(a) (b)

(c)

Figure 1: Task Depiction: (a) Subject is first presented with stationary object outside the FOV and corresponding auditory cue. (b) Subject must localize the object by bringing the object into FOV. (c) Subject must respond by clicking the mechanical clicker.

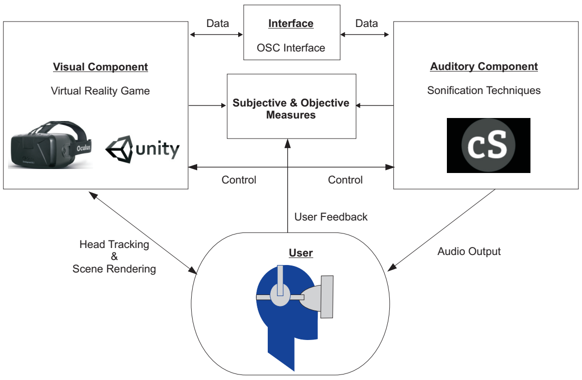


Figure 2: Proposed Model

1. *Visual Component:* This component incorporates a game-based VR that facilitates the real world task of stationary object localization in virtual world. The VR game is implemented with the aid of Oculus Rift[1] and Unity3D[2]. Oculus rift is a virtual reality head-mounted display developed by Oculus VR. Unity3D is a cross-platform game creation system developed by Unity Technologies, including a game engine and integrated development environment. Section 3.2 presents the implementation of this component.
2. *Interface:* The relevant data of the virtual world is sent to the Auditory Component via an Open Sound Control (OSC) interface. OSC is a protocol for communication among computers, sound synthesizers, and other multimedia devices that is optimized for modern networking technology[3].
3. *Auditory Component:* The system's Auditory Component accommodates the sonification methods to be evaluated. The sonification module transforms the data into auditory cues using Csound[4]. The design and implementation of the sonification techniques is explained in Section 3.3.
4. *User feedback and Measures:* The test subject interacts with the VR game in order to locate a stationary object using auditory cues. The subject provides responses via a mechanical clicker. During the process of the interaction, a set of objective and subjective measures are recorded. Different sonification algorithms are used to accomplish the task. Hence, a set of measures are available for each sonification algorithm. Using those measures a quantitative and comparative analysis of the sonification methods can be performed. Section 4.1 explains the user interaction with the system.

3.2. VR Game

The Visual Component, of the proposed model, constitutes a VR game. Figure 3 illustrates the architecture of this component. We have used the Unity3D game engine to develop the visual scene. This virtual visual scene is presented to the test subject. The head tracking accomplished by the Oculus Rift tracks the subject placed in virtual space. The Scene Rendering procedure renders a stereoscopic view of the visual scene via the Oculus Rift. Figure 4 shows the visual scene when the stationary object is not in the FOV and Figure 5 shows the visual scene when stationary object is in the FOV.

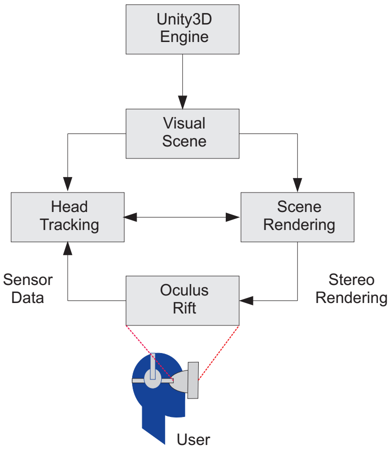


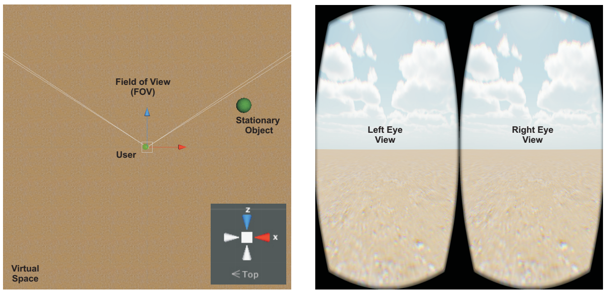
Figure 3: Architecture of the Visual Component.

3.3. Sonification Design

Mapping: In this paper, we use Parameter Mapping Sonification (PMSon) [9] which involves mapping of measured variable or data to value of sound synthesis parameters like frequency, brightness, amplitude. Here the azimuth of the stationary object with respect to the subject in virtual space is used to provide the information of the object's location. This azimuth is mapped to a 180 degree stereo sound field, with the locations behind the listener being mapped in front of them. This mapping results in amplitude

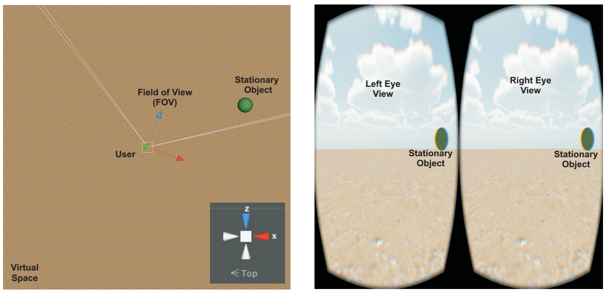
---

[1] `https://www.oculus.com/`
[2] `http://unity3d.com/`
[3] `http://opensoundcontrol.org/`
[4] `http://www.csounds.com/`



(a) VR Game: Top View (b) VR Game: Front View

Figure 4: Visual Scene: Stationary Object Outside FOV



(a) VR Game: Top View (b) VR Game: Front View

Figure 5: Visual Scene: Stationary Object Inside FOV

scale factor for left and right channel of the stereo output which is continuously updated with the head rotation. This results in a continuous panning of the signal into the stereo auditory field based upon the angle between where is subject is looking and the stationary object. Front-Back confusion only occurs when object appears exactly in front or behind the subject and was thus not considered a significant impediment.

Spectrum and Envelope: It is known that the spectral content (spectrum) of a sound source, along with the manner that the content changes over time, is largely responsible for the perceptual quality of timbre [10]. We made a choice of three different spectral types to test the effect of spectral content on the subjects ability to localize the virtual sound source, a pure tone of 1 kHz, white noise (an equal distribution of all frequencies in the spectrum) and a plucked sound which is somewhat in-between the pure tone and the noise. All signals are of one second duration and are attack-decay (AD) enveloped with an attack time of 20ms. These choices are psychoacoustically motivated as humans are known to be sensitive to timbral difference when the attack time of the signal is short [11].

Jitter: We define jitter as a special effect. Humans tend to ascertain the direction of sound source by moving their head. We simulated this effect by oscillating the azimuth of the stationary object by a small azimuth δ. The δ chosen was 5 degrees azimuth at a rate of 15 cycles per second. These values were chosen after making few informal trials. This small movement of the position of the stationary object results in oscillation of the stereo pan variable.

As shown in Figure 6 we combine all the above design elements to create six different sonification methods. The spectrum generated by sine, pluck and noise is AD enveloped. This enveloped signal is then panned with or without jitter. We term the sonification methods as *sine*, *sine + jitter*, *pluck*, *pluck + jitter*, *noise*, *noise + jitter*. We aim to evaluate the effectiveness of signals of different spectrum and also determine if the jitter adds value in the source localization task.

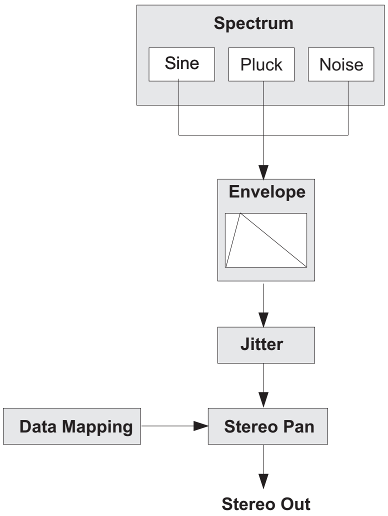


Figure 6: Six sonification methods. Sine, pluck and noise is enveloped and then panned with or without jitter.

4. EXPERIMENTAL SETUP

This section details the experimental setup employed to evaluate the sonification methods introduced in Section 3.3. Section 4.1 illustrates the listening test procedure undertaken. Then, the subjective and objective measures adopted for the listening test is introduced in Section 4.2.

4.1. Test Description

Two test cases were chosen considering that the sonification method could be used in scenarios of total blindness like driving at night or by visualy handicapped. The aim was to determine the difference in response time with the additional visual modality.

1. *Test 1 (Blind):*
   In the first test case the subject was instructed to perform the task of object localization, as described in Section 2, with auditory feedback but without visual feedback. In this case, the visual scene was blacked out and only the auditory cue was presented to the subject.
2. *Test 2 (Sighted):*
   In the second test case the subject had both auditory and visual feedback. The task was the same as for *blind* test case.

Each of the six different sonification methods, as described in Section 3.3, was tested in both the above cases. Each subject had to undertake both test cases. As shown in Figure 7, subjects were presented with stationary objects in the described virtual space at eight different fixed locations for each sonification method for each of the two test cases. A single instance is referred to as a trial, where the stationary object was presented at a single fixed location. In all, each sonification method was tested for eight different locations and for both the cases of *blind* and *sighted* for a total of 96 trials per subject.

The entire listening test was conducted in three sessions. The first was a training session in which the subject became familiar with the use of the Oculus Rift, VR Game and sonification methods. The second and third sessions encompassed the 96 trials. The order of presentation of trials was randomized across all the participants. In order to ensure that there is no motion sickness due to the Oculus Rift, each trial was limited to a maximum duration of 20s and each session took no longer than 15 minutes.

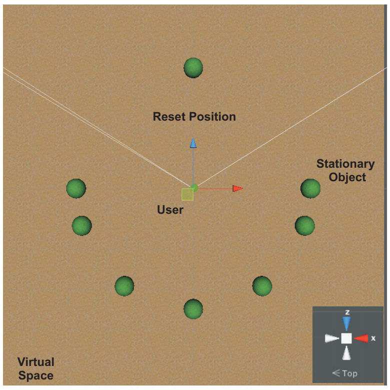


Figure 7: Stationary objects at 8 unique fixed positions for each sonification method for each subject.

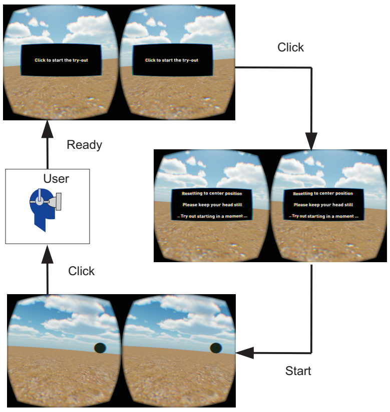


Figure 8: Single trial as seen by the subject. When ready, subject clicks to reset the Oculus Rift orientation to the reset position. Then the test subject is presented with the scene and auditory cue. The subjects localize the object and click to respond.

A single trial as seen by a test subject is depicted in the Figure 8. The subject clicks the mechanical clicker when ready. The position of the subject in virtual space is first reset to a fixed reference position. Once reset, the subject is presented with the scene and auditory cue. In the *blind* test case the subject is instructed to achieve the task of localizing the stationary object to the center of the FOV and then respond using a mechanical clicker. While in the sighted test case, the subject is instructed to bring the object inside the FOV and respond using the clicker as soon as the subject has seen the object. The processing of click sound to measure the response time is described in Section 4.2.2. When the test system detects a click, the trial ends and evaluation measures are recorded. This process is repeated for all sessions.

Thirteen participants took part in the experiment. The average age of the participants was 30.23 years (standard deviation of 11.12 years). The group included people of German, Indian, Spanish, Australian and Chinese nationality. Nine males and four females took part. Some test-oriented questions were asked. It was found that six participants had background in music or audio processing. Four of them had background in audio processing, two in both audio processing and one in music. The remaining seven did not have music or audio processing knowledge. With regards to experience with Oculus Rift, only one out of 13 participants had previously used Oculus Rift. With respect to gaming expertise, seven were amateur, four rated themselves as intermediate and two as experts. With respect to listening test expertise nine were of beginner level and four were expert.

4.2. Evaluation Measures

The sonification methods introduced in 3.3 were evaluated by a listening test in which both subjective and objective measures were recorded.

4.2.1. *Subjective Measures*

We asked the subject to rate each of the sonification method on a 5 point Likert categorical scale [12] according to the agreement with the following statements:

1. ”The mapping of location to sound is intuitive.” This question was asked collectively for all the sonification methods because the mapping remains the same for all methods.
2. ”The sound is pleasant.” This question was asked for each sonifcation method.

4.2.2. *Objective Measures*

In order to objectively evaluate how quickly the subject responds to different sonification method, the response time of subjects was measured. The subject’s response time was recorded using a system of mechanical clicker, audio recorder and onset detector. We call this time *responseTime*. The click has dual function in the course of the experiment; to both start/stop the trial and to record the response time of the subject.

Figure 9 is a block diagram of the system used to record the sound of the clicker and to determine the response time. The equipment used was an RME Fireface 400 – a firewire audio and midi interface developed by RME[5]. We used this to record both the input signal from the microphone and the output signal from the Auditory Component simultaneously. The stereo output from Auditory Component was sent to headphones as well as re–routed using external cables back to the Fireface interface. Hence, we have a 3-channel audio setup, one channel with click sound and the second and third with left and right audio from Auditory Component. Since we are using a standalone hardware interface the audio signal sent to headphones and re–routed reached the headphones and Fireface input channel at the same time. This ensured

---

[5] `http://www.rme-audio.de/en/company.php`

that there was no processing delay when calculating the reaction time of the subject. It is crucial to avoid system delays because the sonification methods are compared in terms of reaction times. The onset detector is applied to the recorded signal in run-time. We used Musical Onset Database And Library (Modal)[6] for real-time onset detection. This real-time onset detection is carried out with linear prediction and sinusoidal modeling, based on the work of Glover et al [13].

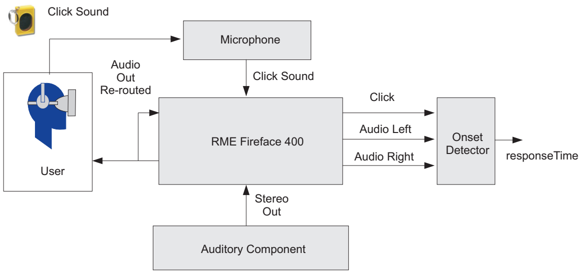


Figure 9: System to measure subject response time.

This onset detector aids in precise detection of response time and is illustrated in the following steps:

1. Subject clicks to start the trial. At this point the 2nd and 3rd channel contains no signal since the task is not yet in action and no sonification is in progress. A simple predefined threshold is measured in the 1st channel to detect the click. Once detected, the trial begins
2. The trial starts and all the three channels are recorded simultaneously. The subject tries to find the object.
3. Subject clicks once more when task of object localization is complete. Again, a simple predefined threshold is measured in the 1st channel to detect the click. Once detected the trial stops.
4. Finally, before the next trial is scheduled, the onset detector is applied on the signal in all the 3 channels. The onset detector detects click onsets on 1st channel corresponding to start and end of the trial. It also detects the onsets in 2nd and 3rd channel corresponding to the start of sonification. Thus, *responseTime* is the time difference between the first onset among 2nd and 3rd channel and the onset corresponding to second click.

Because all the audio signals are recorded in the same audio interface, this method ascertains the precision of the response time. Because we take the first onset among those channels, it also ascertains precision in the case where the object is exactly right or left of the subject. The audio was recorded at 44.1 kHz and also the onset detection was essentially performed after the trial using the recorded signal. Hence, there was no processing lag introduced in response time measurement. Figure 10 shows the audio recording setup of a listening test trial.

It was initially thought that a simpler method such as keyboard/mouse input could be used to record the response time. But a performance profile of the implemented system of auditory virtual reality game showed the Visual Component consumed a maximum of 66ms CPU usage and 16ms of GPU usage to run a single routine. Furthermore, this value was not constant over time which was unacceptable when accurate measurements were required. The OSC message exchange took only 6μs per data exchange; Hence negligible. The time taken by Auditory Component i.e. Csound was also negligible. However, a non-constant delay of the order of 66ms was unacceptable for comparing the response time between different sonification methods. For example, results discussed in Section 5 show that the difference in response time is of the order of 100ms to 150ms. Hence a 66ms routine would result in an erroneous measurement. Therefore, a decision was made to use a system of mechanical clicker, audio recorder and onset detector to measure the response time.

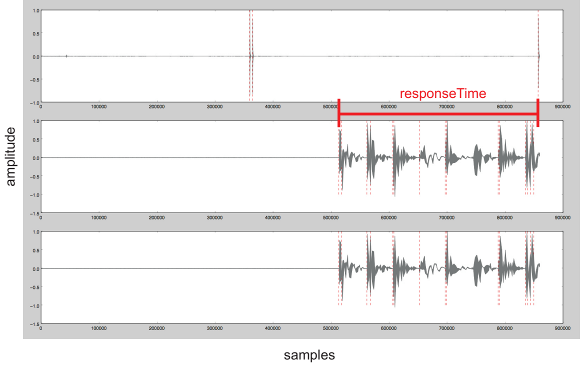


Figure 10: Onset Samples: Top – Click signal, Center – Sonification signal left channel, Bottom – Sonification signal right channel; Dotted Red Line – Onsets detected.

5. RESULTS

The sonification methods were evaluated for the measures defined in Section 4.2. The main aim was to find out if there was any significant difference in response time between the six sonification methods. The response times were then used to determine the effect of the auditory feedback on the task. It was expected that the *sighted* test case would result in faster response time than in the *blind* test case.

Figure 11 shows a boxplot of the reaction time with and without visual feedback (namely, Sighted/Blind on the x-axis) and the sonification method factors. Mean values are shown as black circles on the box plot for each of the conditions. We observe that the mean reaction time is statistically lower in *sighted* test case when compared to *blind* test case for each sonification method except for for *noise + jitter* as seen in multicomparision. Also, the Interquartile Range (IQR) for the test case *sighted*, has less spread than for the test case *blind*, for each of the sonification method. This is probably because an additional visual modality provides more information in *sighted* test case. One other important observation is that the IQR of *noise* is smaller when compared to other methods, implying that the *noise* sonification method was more predictable and consistent over other methods.

---

[6] `https://github.com/johnglover/modal`; A cross-platform library for musical onset detection, written in C++ and Python. It is provided under the terms of the GNU General Public License

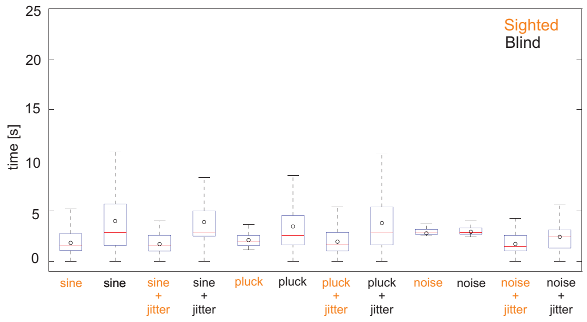


Figure 11: Boxplot describing the effect of the sonification methods and visual feedback on the task of object localization.

5.1. Global effect of the visual feedback and sonification methods

In order to assess the effects of the sonification methods and the complexity of the visual search experiments, a 3–way ANOVA analysis of the reaction times has been performed. The sonification methods and the use of visual feedback were fixed factors, whereas subjects were set as a random factor for the statistical analysis.

First we checked how much the performance varied between subjects. A multiple comparison was performed to check which pairs of means are significantly different over the different factors. It results in the estimated inter mean group difference and the confidence interval for the compared group. Figure 12 shows a multicomparison of the subjects over all the sonification method and test cases. It can be seen that subjects 6, 7, 8, and 10 performed worse than at least one of the best group of subjects. Therefore we decided to eliminate these subjects and evaluate the sonification methods based on the remaining subjects and we call them expert subjects.

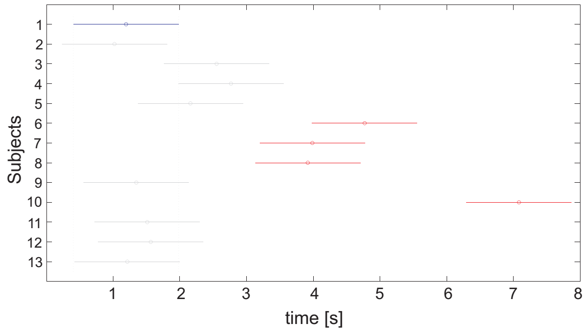


Figure 12: Multicomparision of subjects against all sonifcation methods and test cases. Subjects 6, 7, 8 and 10 perform worse than at least one subject from our expert group.

Next, we proceeded with the multicomparision of sonification methods against the *blind/sighted* test cases and Figure 13 shows these results. It can be observed that the mean response time in test case *sighted* is statistically significant than *blind* test case for each sonification method except for *noise + jitter*. Interestingly, the *noise + jitter* in blind case performed statistically better than the *noise* in *sighted* case. Also, we see that when jitter is added, the sonification methods tended to perform better than its counterpart except for *pluck* in *sighted* test case. However, the differences are not significant.

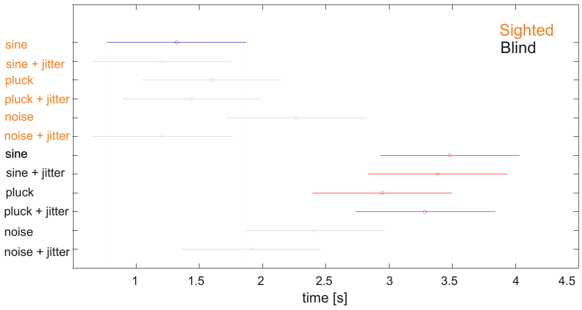


Figure 13: Multicomparision of sonification methods against all the two test cases. Top six is for the *sighted* case and bottom six are for the *blind* case.

Results of the ANOVA analysis show that the sonification methods did not have a statistically significant effect on the average reaction time ($p = 0.1533 > 0.05$), while visual feedback had a statistically significant effect on the average reaction time ($p = 0.0112 < 0.05$). We now analyze each of the individual factors, visual feedback and sonification methods, in more detail.

5.2. Local effect of sonification methods for each test case

We then analyzed the sonification methods as a factor and compared reaction times for both the test cases (*sighted* and *blind*). For this, a 2-way ANOVA of the reaction times for all sonification methods in the two test cases was performed. Here, the sonification method is a fixed factor and subjects are a random factor.

Although we expected noise, due to its wider spectrum, to result in faster response, we can observe that *noise* performed statistically worse than all other methods in *sighted* test case. There are no statistically significant differences among the remaining 5 methods, as shown in the multicomparision plot of Figure 14. Signal with jitter tended to perform statistically better than its counterpart but it is not significant in *sighted* test case.

As shown in Figure 15, for the *blind* case we see that *noise + jitter* performed statistically worse than all the other methods but *noise* and *pluck* . The performance of jitter is inconsistent in the *blind* case.

Even though *noise + jitter* has statistically better objective performance, as we will see in Section 5.3, it had had lower subjective rating. Also, *pluck + jitter*, which is rated subjectively higher by subjects, performed statistically worse in *sighted* test case.

5.3. Subjective Measures

Statement 1 – The mapping of location to sound is intuitive:
Figure 16 shows the responses of the participants. 7 out of the 13 participants strongly agree to this, while none disagreed.

Statement 2 – The sound is pleasant:
This was asked for every sonification method in both the test cases. The outcome is as shown in Figure 17. *Noise + Jitter* has attracted more negative reviews compared to other methods. In all other cases the methods with jitter seemed to be preferred over its counterpart. In general, noise was also rated lower compared to sine and pluck by the subjects.

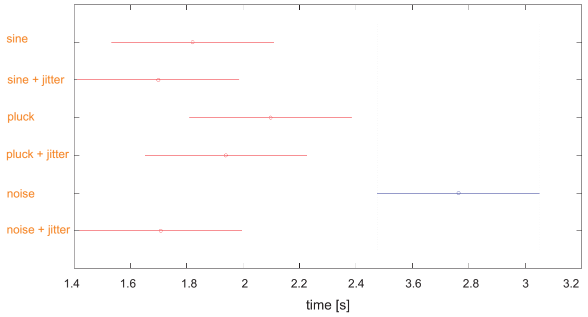


Figure 14: Multicomparision of sonification methods in the *sighted* test cases.

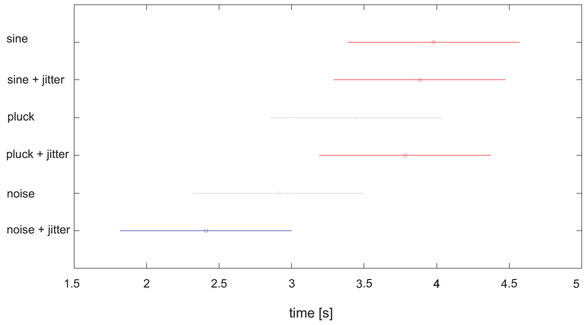


Figure 15: Multicomparision of sonification methods in the *blind* test cases.

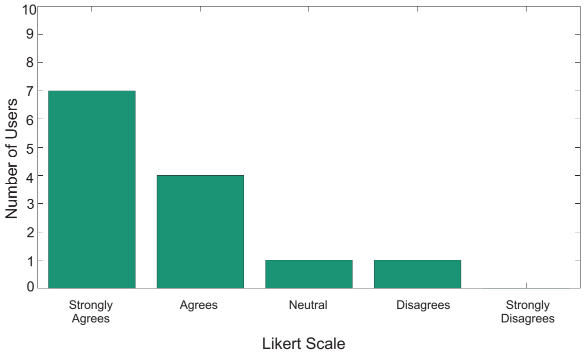


Figure 16: Likert scale for: *Mapping of the location to sound is intuitive.*

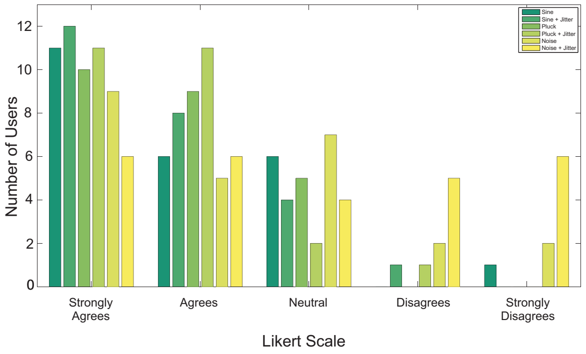


Figure 17: Likert scale for: *Sound is pleasant.*

6. CONCLUSIONS AND DISCUSSION

This paper describes a virtual reality platform as reliable and scalable experimental setup for evaluation of sonification methods. We have defined the task of source localization in virtual space and developed a system to facilitate the same. Further, perceptually motivated sonification methods were designed and implemented using PMSon method. We also defined response time as objective measure and implemented a robust and low latency system to determine the response time of the subject. Subsequently, a formal listening test was performed to subjectively evaluate the performance, aesthetics and intuitiveness of the designed sonification methods. Finally, the subjective and objective measures were statistically analyzed to achieve a scientific comparison of the sonification methods. The results show that *noise + jitter* performed significantly better than other methods in *blind* test case but was rated subjectively low by test subjects. Although we expected noise, due to its wider spectrum, to be statistically better than other methods, in sighted case *noise* performed statistically worse than other methods. The *noise* method tended to be more robust in both *blind* and sighted test cases in terms of IQR, but it was rated as less pleasant than other methods. Another clear result is that, with the addition of visual modality mean response times were significantly shorter except for *noise* in *sighted* test case. With addition of jitter better response times were recorded in *sighted* test case but not statistically significant than its counterpart, while in *blind* test case it was inconsistent.

Real world applications require more complex tasks such as localizing moving objects, navigation. As an extension to this work, we could add to the complexity of the system to accommodate such complex tasks. Further, other sound parameters like loudness, reverberation, 3D binaural could be explored to facilitate such complex tasks. This system could also be employed to evaluate parameter optimization in sonification. The use of a game-based virtual reality for scientific experiment has been exhibited. The methodology could be adapted in similarly motivated work like audio listening test experiments, sonification for medical and industrial applications. and might be useful in blind navigation, in-vehicle sonification for driver assistance etc. experiments.

This research validates the use of a virtual reality system in sonification evaluation. By being both scalable and reproducible, it makes a contribution to reproducible research in sonification.

7. REFERENCES

[1] A. D. N. Edwards, "Auditory display in assistive technology," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 17, pp. 431–453. [Online]. Available: http://sonification.de/handbook/chapters/chapter17/

[2] M. Bujacz, P. Skulimowski, and P. Strumillo, "Naviton—a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired," *J. Audio Eng. Soc*, vol. 60, no. 9, pp. 696–708, 2012.

[3] D. El-Shimy, F. Grond, A. Olmos, and J. Cooperstock, "Eyes-free environmental awareness for navigation," *Journal on Multimodal User Interfaces*, vol. 5, pp. 131–141, 2012.

[4] D. Storek, F. Rund, T. Barath, and S. Vitek, "Virtual auditory space for visually impaired - methods for testing virtual sound source localization," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2013, pp. 33–36.

[5] T. Hermann, "Taxonomy and definitions for sonification and auditory display," in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, P. Susini and O. Warusfel, Eds. IRCAM, 2008.

[6] T. L. Bonebright and J. H. Flowers, "Evaluation of auditory display," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 6, pp. 111–144. [Online]. Available: http://sonification.de/handbook/chapters/chapter6/

[7] N. Degara, F. Nagel, and T. Hermann, "SonEX: an evaluation exchange framework for reproducible sonification," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2013, pp. 167–174.

[8] J. Snchez, J. Snchez, J. Snchez, M. Lumbreras, M. Lumbreras, M. Lumbreras, J. P. Silva, J. P. Silva, and J. P. Silva, "Virtual reality and learning: Trends and issues."

[9] F. Grond and J. Berger, "Parameter mapping sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin, Germany: Logos Publishing House, 2011, ch. 15, pp. 363–397. [Online]. Available: http://sonification.de/handbook/chapters/chapter15/

[10] J. Grey, *An Exploration of Musical Timbre Using Computer-based Techniques*. Department of Psychology, Stanford University., 1975. [Online]. Available: http://books.google.de/books?id=yQ5HAAAAIAAJ

[11] E. Zwicker and H. H. Fastl, *Psychoacoustics : facts and models*, ser. Springer series in information sciences. New York: Springer Verlag, 1998. [Online]. Available: http://opac.inria.fr/record=b1094760

[12] K. A. Dukes, "Wiley statsref: Statistics reference online," *Wiley Online Library*, 29 SEP 2014.

[13] V. L. John Glover and J. Timoney, "Real-time detection of musical onsets with linear prediction and sinusoidal modeling," *EURASIP Journal on Advances in Signal Processing*, vol. Volume 2011, September, no. 68, 2011.