INVESTIGATING THE USE OF VIRTUAL REALITY TO SOLVE THE UNDERLYING PROBLEMS WITH THE 3D STAGE PARADIGM

William Gale
School of Computing and Engineering
University of Huddersfield
u1457211@unimail.hud.ac.uk

Jonathan Wakefield
School of Computing and Engineering
University of Huddersfield
j.p.wakefield@hud.ac.uk

ABSTRACT

3D Stage Paradigm (SP) interfaces have been shown to outperform traditional DAWs in speed, mix overview and satisfaction. However, SP interfaces raise problems of their own including clutter, object occlusion, depth perception, object interaction, exit error and gorilla arm. Building on from previous research this project implemented a 3D SP interface in Virtual Reality (VR) to try to address these problems. A formal usability evaluation focused on efficiency, effectiveness and satisfaction was conducted. Three VR and desktop interfaces were created for two micro task and one macro task-based tests. Results showed VR was as efficient as desktop but slightly less effective. Furthermore, there was a significant preference towards VR. Results indicated clutter, object occlusion, and exit error are not improved. However, gorilla arm, and depth perception appear to improve in VR.

1. INTRODUCTION

Mixing desk interfaces mainly use the Channel Strip Paradigm (CSP), with arrays of faders, buttons and knobs controlling mix parameters with one to one mappings. CSP has been implemented on analogue and digital mixing desks, Digital Audio Workstations (DAW), web-based mixers and iPads. Many attempts have been made to design alternative interfaces, with one prominent example being the Stage Paradigm (SP). Studies have shown SP can outperform CSP at certain tasks [1]. However, existing SP implementations present a number of issues that need to be addressed.

2. BACKGROUND

Stage Paradigm is a popular alternative mixing interface design, first introduced by Gibson [2]. It loosely emulates how humans naturally perceive audio from real world objects [3]. Usually SP consists of visual *audio objects* e.g. spheres. Different properties of the sphere denote different mix parameters. A typical 2D implementation might use the X and Y screen coordinates, respectively, to represent the pan and level of the audio. Many examples of SP have been trialled, for example, Gelineck et al. [4] who created a range of interfaces based on a touch screen and physical *smart tangible* controllers, with the previously described mapping of pan and level. Participants liked SP in general and having physical controllers allowed very accurate control, whereas the touch screen was deemed less accurate due to *exit error*. *Exit error* occurs due to micro-movements of the hand as the user tries to let objects go resulting in inaccurate placement. *Visual clutter*, i.e. too many objects potentially overlapping and/or occluding each other on the screen, was also an issue for all interfaces. A potential solution for clutter is to add a third dimension (3D), and adopt a *depth mixing* approach [5] using the Z axis. Wakefield et al. used this approach in their LAMI interface [6]. This interface utilized a Leap Motion for control and a pseudo-3D visualisation on a 2D computer screen. *Exit error* was found to be an issue. Furthermore, due to the Leap Motion, users complained of *gorilla arm*. *Gorilla arm* occurs when users have to keep their arm elevated to use a controller [7]. Despite these problems, user keyword selection suggested LAMI was more *fun* than the benchmark traditional DAW interface. Several other SP implementations have been created that faced similar problems [3, 4, 8, 9].Furthermore, using pseudo-3D visualisations makes depth perception difficult for users which can make interaction with visual objects difficult. This research aims to investigate whether the use of Virtual Reality (VR) can address these problems.

3. THE VR INTERFACE

Figure 1 shows the VR interface developed for this experiment. Spheres visually represent audio tracks. The name of the audio track is displayed upon the sphere. The X position of the sphere represents the corresponding audio track's pan, whilst the Z position represents its level. The interface was developed in Unity [10] and a Gear VR headset was used [11]. Seven controls were used on the Gear VR controller, these will be referred to as *Primary*, *Secondary*, *Scroll*, *Button1* to *Button4* and can be seen as labels on the images on the right hand side of *Figure 2*.

To move a sphere, users must highlight it by pointing the controller's laser beam at the sphere, and then select it by holding the *Primary* button and moving the controller to point to the desired position. The *Scroll* touchpad is used to move the sphere in the Z axis. Audio tracks can be soloed by highlighting a sphere and pressing *Button3* or muted by pressing *Button1*. Pressing *Button1* or *Button3*, without highlighting a sphere unmutes/unsolos everything. Highlighting and pressing *Button2* enables delay and reverb sends to be shown as small white spheres. These are controlled by highlighting them, holding *Primary* and moving them left/right with the visualisation moving in an arc. Highlighting an sphere and pressing *Button4* shows the EQ visualisation. This allows control of a lowpass, highpass and peaking filter. Highlighting and selecting the wider top/bottom of the EQ visualisation allows control of the frequency of a highpass/lowpass filter. The small sphere to the right side of the EQ visualisation allows control of a peaking filter. Moving this up/down controls the centre frequency, left/right controls cut/boost and the *Scroll* touchpad controls Q.

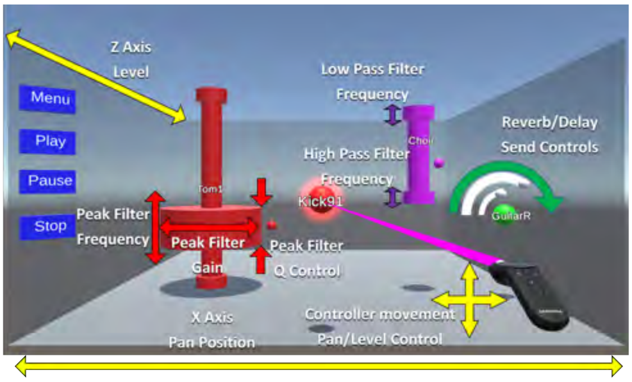

Figure 1: Overview of the interface

Figure 2: Interface Controls

A PC version of the VR interface was developed as a benchmark. This PC interface provided a pseudo-3D visualisation on the computer screen and used the control mapping shown on the left hand side of *Figure 2* based on a computer keyboard and mouse.

4. EVALUATION

A formal usability evaluation of the interface was conducted. This consisted of three tests, namely, Clutter Test, Accuracy Test and Full Mix Test. The first two tests were micro task-based tests while the last was a macro task-based test. Each test evaluated the VR interface and did the equivalent test on the PC version to provide a benchmark. Ten Music technology students participated in each test. These tests aimed to measure efficiency, effectiveness and satisfaction.

4.1. Clutter Test

In each iteration of the Clutter Test, one of four randomly selected mix sessions was presented to the user. These sessions consisted of 6, 12, 24 or 32 audio tracks. Users were asked to select a target track from a visualisation of a mix session. The name of the target track was shown at the top of the interface. Subjects were asked click on the target track as quickly as possible. Green or red text appeared at the top of the visualisation, denoting a correct or incorrect selection. When the correct track was selected, another mix session and target were randomly chosen. The random selection was controlled so that each mix session appeared 5 times. Subjects had to find 20 tracks in total. Unlike Dewey et al. [12], each mix session had predefined track locations to make the test more realistic. Similarly, track names were assigned like a real mix scenario. Tracks with L or R in their name were placed to the left or right respectively. Furthermore, tracks such as Kick, Bass or Snare were placed in the centre as this is a common position for these instruments. Subjects could practice using each interface before being tested. This test was designed to measure efficiency by measuring the time taken to select each track.

4.2. Accuracy Test

In this test the user was asked to move a white sphere to the purple target sphere as quickly and accurately as possible. A *Next* button was provided for the user to indicate they were happy with the placement of the white sphere which then moved them on to the next iteration of the test. Timing began when the subject first clicked on the white sphere until when they released the sphere. Moving the sphere again resumed timing. This ensured only time spent moving the sphere was recorded. Subjects could practice with the interface before the test. Subjects were presented with 10 randomly positioned targets in total. This test was designed to measure effectiveness by measuring the distance from the user positioned sphere to the target sphere. Efficiency was additionally measured using the time taken to position each white sphere.

4.3. Full Mix Test

Subjects were asked to mix a 21 track song using the full feature set of the interface. Subjects could spend as long as they wanted using the interface. Once they were happy with their mix, they completed a short questionnaire. This consisted of picking five words describing their experience with the interface. A reduced selection of words from the Microsoft Desirability Toolkit [13] were used. Subjects were also asked to score the interface out of ten. This test was designed to measure the satisfaction between the desktop and VR interfaces.

Verbal comments were also elicited from the test subjects.

5. DISCUSSION OF RESULTS

5.1. Clutter Test

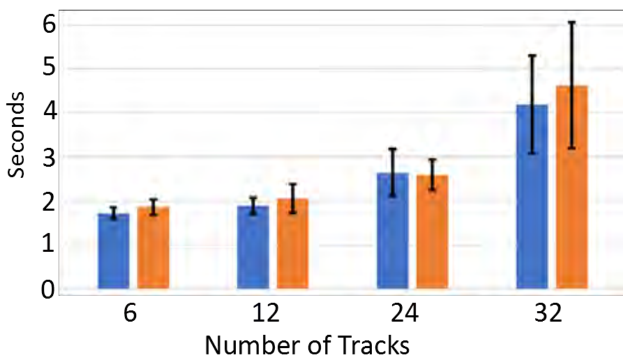

Figure 3: Average time to find the target track for each mix session with 95% confidence intervals. Blue bars represents PC interface. Orange bars represents VR interface.

Figure 3 shows the average time taken for all subjects to find the target track for each mix session. A univariate analysis was performed which showed that track count had significant effect on time ($p = 0.000$, $F = 17.532$) regardless of whether it was the VR or PC interface. However, the results showed that there was no significant effect on time ($p > 0.505$, $F = 0.445$) between the VR and PC interface. This suggests that using VR does not assist in addressing clutter. Some subjects complained about objects occluding each other which suggests that VR does not improve issues with object occlusion.

Results were further analysed by splitting tracks into easy to find and hard to find track names. For example, Guitar L is easy to find as the name hints it will be positioned to the left. Kick is also an easy to find track name as generally this track is panned to the center. Triangle is hard to find as the name does not hint at the pan position, nor does it have a standard pan position. Results showed a significant effect of easy/hard to find track names on the time ($p < 0.002$, $F = 9.964$) regardless of whether it was the VR or PC interface.

5.2. Accuracy Test

The left hand side of *Figure 4* shows the average times taken to move the sphere to the purple target sphere. The right hand side shows the average accuracy of the sphere placement.

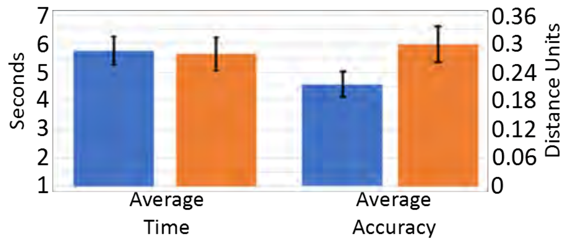

Figure 4: Average time and accuracy of sphere placement with 95 % confidence intervals. Blue bars represents PC interface. Orange bars represents VR interface.

A one-way ANOVA test was performed and results showed that there was no significant effect on time taken ($p = 0.774$, $F = 0.83$) regardless of whether it was the VR or PC interface. However, there was a significant effect on the accuracy ($p = 0.001$, $F = 12.2$) with the PC interface being more accurate. Despite this, the difference between interface accuracy was only approximately 0.1 Unity units. For perspective, the spheres in the visualisation are 1 Unity unit wide. It is questionable whether this difference would be audible.

Some subjects commented that the audio spheres would not stay where they placed them in the VR interface. This suggests that this VR interface does not improve exit error. Some subjects commented that they found depth perception easier when using the VR interface, however, this did not lead to improved performance in this test. It is difficult to comment on object interaction as the input mechanisms for each interface were different.

5.3. Full Mix Test

Figure 5: Word Clouds

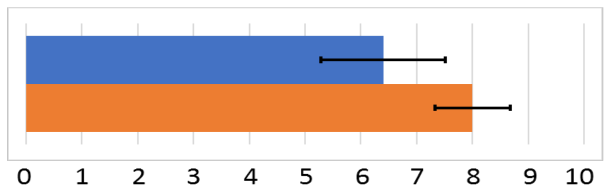

Figure 6: Average preference scores. Blue bar represents PC interface. Orange bar represents VR interface.

Figure 5 shows word-clouds of the words subjects selected in the questionnaire. Word size represents the frequency that words were chosen. It is clear subjects found the VR interface more *fun, engaging and exciting*. However the PC interface was more *stable, effective and simplistic*. This could be because subjects are more familiar with PC interfaces due to using them everyday. Whereas subjects had little to no experience with VR and it is novel to them. This is backed up by a comment from a subject stating "desktop using mouse is usual, VR is completely new". All subjects sounded more excited when using the VR interface and were all eager to use VR as quickly as possible. Due to the familiarity with PC interfaces there is less of a learning curve. This could be why words like *straight-forward*, *easy-to-use* and *familiar* were chosen.

Figure 6 shows the average preference scores for each interface with 95% confidence intervals. Analysis using a one-way ANOVA shows a significant preference for the VR interface ($p = 0.035$, $F = 5.189$).

There were no complaints from test subjects of gorilla arm. The authors believe this is due to the design of the Gear VR controller.

5.4. Test Limitations

This test was limited to only one VR headset and associated controller. Using a higher quality headset like the HTC Vive [14] could give different results. This has more accurate tracking than Gear VR headset which could reduce the accuracy difference between PC and VR. Whilst the Gear VR headset did not even allow forward/backward movement of the head, the HTC Vive headset facilitates this, and furthermore, allows the user to walk around the room. This could potentially help visual clutter and object occlusion.

6. CONCLUSION

This project tested whether VR could address problems with existing SP implementations. Formal usability evaluations were performed on ten music technology students. Three tests were performed measuring efficiency, effectiveness and user satisfaction. Two tests were micro task based and one was macro task based. Most subjects found the VR interface to be fun, while the PC interface was more stable with a significant preference for VR. From the results, VR potentially helps with *depth perception* and *gorilla arm*. However, *clutter*, *object occlusion* and *exit error* appear to be still problematic.

7. REFERENCES

[1] C. Dewey and J. P. Wakefield, Novel Designs for the Audio Mixing Interface Based on Data Visualisation First Principles, presented at the 140th International AES Convention, Paris, France, 2016.

[2] D. Gibson, U.S. Patent No. 6,898,291. Washington, DC: U.S. Patent and Trademark Office. 2005.

[3] J. Ratcliffe, Hand motion-controlled audio mixing interface. Proc. of New Interfaces for Musical Expression (NIME), London, England, 2014, (pp.136-139).

[4] S. Gelineck, D. Overholt, M. Bchert, and J. Andersen, Towards an Interface for Music Mixing based on Smart Tangibles and Multitouch, presented at the NIME, Daejeon, Korea, 2013.

[5] Carrascal, J. P., and Jord, S. Multitouch Interface for Audio Mixing. Presented at the NIME, Oslo, Norway 2011 (pp. 100-103).

[6] J. Wakefield, C. Dewey, and W. Gale, LAMI: A Gesturally Controlled Three-Dimensional Stage Leap (Motion-Based) Audio Mixing Interface, in Audio Engineering Society Convention 142, Berlin, Germany, 2017.

[7] S. Boring, M. Jurmu, and A. Butz, Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays, In Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7, Melbourne, Australia, 2009, (pp. 161-168).

[8] S. Gelineck and D. Korsgaard, An Exploratory Evaluation of User Interfaces for 3D Audio Mixing, presented at the Audio Engineering Society Convention 138, Warsaw, Poland, 2015.

[9] S. Gelineck and D. Overholt, Haptic and Visual Feedback in 3D Audio Mixing Interfaces, in Proceedings of the Audio Mostly 2015 on Interaction With Sound, New York, NY, USA, 2015, pp. 14:114:6.

[10] Unity, Unity. [Online]. Available: https://unity3d.com. [Accessed: 12-Aug-2018].

[11] Gear VR with Controller, Samsung uk. [Online]. Available: //www.samsung.com/uk/wearables/gear-vr-r325/SM-R325NZVABTU/. [Accessed: 12-Aug-2018].

[12] C. Dewey and J. Wakefield, Formal Usability Evaluation of Audio Track Widget Graphical Representation for Two-Dimensional Stage Audio Mixing Interface, in Audio Engineering Society Convention 142, Berlin, Germany, 2017.

[13] Microsoft Desirability Toolkit Product Reaction Words, Nielsen Norman Group. [Online]. Available: https://www.nngroup.com/articles/desirability-reaction-words/. [Accessed: 29-Apr-2018].

[14] VIVE — Discover Virtual Reality Beyond Imagination. [Online]. Available: https://www.vive.com/uk/. [Accessed: 05-Dec-2017].