Conduction an Orchestra in Virtual Reality

Arnór Rafn Gíslason and **Ármann Ari Árnason** and **Hjalti Steinar Sigurbjörnsson**
Department of Computer Science
Reykjavik University
Menntavegi 1, 101 Reykjavik, Iceland
{arnorrg15 | armann15 | hjalti15} @ ru.is

Abstract

This paper explores how conducting an orchestra can be translated into the world of virtual reality. How a conductor's movements and gestures can be interpreted by modern virtual reality technology and How a conductor's movements and gestures can be interpreted by modern virtual reality technology but still be intuitive and engaging to novice users. Each motion and gesture should be easily performable by itself but when doing two or more simultaneously, they should be less intuitive. We asked novice users to test a virtual reality video game that we created and interviewed them in 3 different sessions. Previously we had defined 5 core features to be tested. Resulting from the tests and interviews we got an idea for how engaging and intuitive a virtual reality orchestra conducting experience can really be.

Introduction

Virtual Reality (VR) equipment has dramatically increased the dimensions of inputs that can be used and captured during real-time simulations, including the position and orientation of the head, hands and arms, as well as a variety of buttons. Given these inputs, one of the primary challenges facing the development of VR applications is one of user interface design: for any desired interaction, which input dimensions should be preferred, and which values should have which effects on the simulation? The goal of this project is to investigate these questions in the context of conducting a small musical orchestra as well as getting insight into how an orchestra conductor influences his orchestra, and in what way his actions could be interpreted in virtual reality.

A conductor may to some extent be portrayed as the director of the music. Just like actors in films or on stage, the musicians know their role and can play their part perfectly without any interference from the conductor. One of the roles of a conductor in a musical orchestra performance, is to put his interpretation of the score on the performance and capture the mood the original composer wanted people to experience. The conductor also must make sure everyone is on the same tempo and ensures correct entries by various members of the orchestra.

A big part of the gestures a conductor displays during a performance is his emotional interpretation of the song. Both hand and facial expressions are performed with emotion to instruct the musicians how they should be playing. These emotional gestures can't be captured with modern VR equipment and the changes that the conductor is implying are so detailed that it is almost impossible given the tools that are available to us.

Microinteractions

When a user is immersed in VR with all its available inputs it becomes important that the experience is approachable and intuitive. This is where microinteractions become increasingly important, especially when the task is conducting an orchestra. When an action can be performed alongside a primary task, it is called a microinteraction (Wolf, Rohs, Naumann, & Müller 2011). An example of this would be shifting gears on a stick shift when steering would be the primary task.

A player stands before a set of instruments of which he can manipulate the sound coming from each one of them. Let's say the user wants to keep a steady tempo for the orchestra with one hand and simultaneously increase the volume of a single instrument with the other, it's important that either one of these interactions is simple enough to be considered a microinteraction. That one task can be performed along with the other without each of them needing all our complete cognitive effort. If both interactions separately require the users' full attention, it becomes impossible to perform the given tasks simultaneously. Not to say that any user should be able to execute these interactions together without practice, rather that it is possible and as effortless as can be.

In the previous example, where the user is finding the right tempo with one hand and manipulating the volume of an instrument with the other, we expect that the volume manipulation will require more focus on behalf of the user than finding the right tempo. We arrive at this speculation because we know that for the user to control a volume of a

specific instrument, he must somehow select an instrument from the orchestra. This requires the use of at least two inputs, the orientation of the controller and either a gesture made with the same controller or the press of a button. Once the instrument is selected, another input must be captured to increase or decrease the volume, this can again

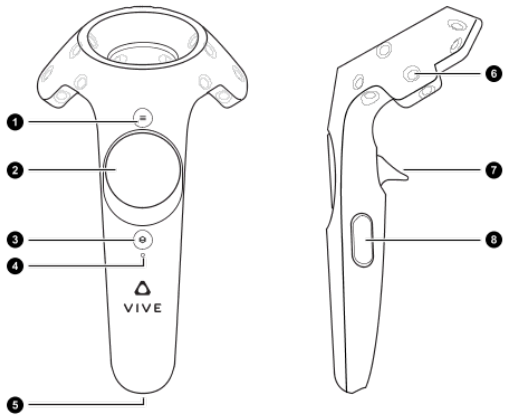

Figure 1: HTC Vive controller

be either a gesture or the press of a button.

In our initial implementation, the user points at an instrument, with the help of a laser cursor, and holds down the trigger button (button 7 on figure 1) while raising/lowering his hand to either increase or decrease the volume. While this is a straightforward and easy to accomplish act, it takes up a lot of the users' cognitive effort making it difficult to enact alongside another operation that is not considered to be a microinteraction.

It is therefore important to keep the tempo manipulation simple enough to be a microinteraction. That way the user can conduct the orchestra without loosing control of the tempo in which the instrument is playing.

Audio and Wwise

At the early stages of our project we used simple wave files as we were figuring out how to implement certain gestures and events in Unity for the HTC Vive headset. After we had gained knowledge on how to work with Unity and Vive we started using Wwise for event driven audio manipulation. Wwise offers real-time parameter control (RTPC) so that certain changes can be applied to the audio at runtime.

Volume manipulation and the pitch of the song is the only thing we could control using the wave files, so for this project another solution had to be used and MIDI files were chosen. MIDI files give a lot more control over the audio. Each individual note cannot be accessed through Unity, however, it can be brought into any digital audio workstation and fine-tuned for the effect desired for this project. MIDI files also allow the playback speed of the track to be changed without affecting the pitch.

Core Features

Hand coordination plays a big part in this simulation as a conductor uses both his left and right hand simultaneously. Because of this, the questions are: "Can a novice user, with little or no hand coordination, operate two controllers (left and right) concurrently and is it possible to create an intuitive and engaging experience, while still offering a variety of ways to influence the orchestra?"

Manipulating Volume

We had three users come in and test the first version of our demo. In it, the user stands in front of a set of instruments that are playing a song. Laser pointers were pointing from the top of both controllers to indicate where they were pointing. To interact with an instrument, the player had to point at it with either controller, click a trigger button on that same controller and then raise or lower that controller to manipulate the volume of the instrument.

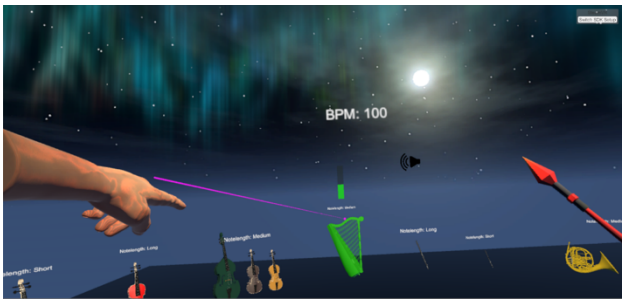

Figure 2: The laser pointer, volume indicator and bpm indicator

At the time we created the demo, we thought it to be obvious and intuitive to control the volume in that way. But once the users started playing, they immediately tried to change the volume by pointing the laser pointer up or down instead of raising or lowering the controller itself. They did this besides the fact that we had taught them the controls and they had seen the previous user make this same mistake before, without the intended results.

The laser trail was obviously causing this confusion, so we removed it but kept the cursor to indicate where the controller is pointing. We tested this on four users, and it turned out to be successful. The users stopped encountering the problem right away when they saw that the visual feedback was not affected at all and the laser trail was not confusing them.

With that issue resolved, the user still had no way of knowing what the volume was. Adding a volume slider to each instrument, the user could see the change in volume and therefore they would immediately figure out that the

volume was not being affected when performing the wrong gesture. As expected this solved the confusion completely and we did not encounter this issue with other testers.

Visual feedback is crucial for developers when creating any software or video game that tries to be intuitive. Users playing a game they are not familiar with and perhaps don't understand, give another set of eyes on problems that developers can be blind to. In other words, developers can't see the forest for the trees.

Manipulating Tempo

The way a conductor keeps the orchestra together and playing at the right tempo, is by swinging his hands at the correct tempo. We wanted to include this feature in our game. The implementation that came to mind was to have the user press a button on the right controller and since the song that is being conducted is in 4/4, the system counts how long it takes the user to swing his/her right hand back and forth 4 times. We record the time it takes the user to perform these 4 swings and put it in an equation to calculate the BPM (beats per minute) and change the tempo of the song accordingly.

The first problem we encountered with this approach was how to record each stroke. To debug this issue, we logged each time the controller switches direction and set a counter that is incremented at each direction change. The problem with this approach is that this does not recognize the first stroke since it is not a change of direction. To counter this problem, we came up with a solution that the first stroke would always be to the right. Which, in terms of conducting, makes a lot of sense since the conductor tends to wave his hand to the right after the upbeat.

Manipulating Note Length

When conductors conduct an orchestra, they can influence for how long the musician holds each note. To do this, they use either long or short strokes. If the strokes are long the orchestra plays softer and holds notes longer, and if the strokes are short the orchestra plays with more power and shorter notes. To add this functionality to our game, we want the user to be able to manipulate the note length of individual instruments and as well as the note length of all instruments at once. We added an indicator under each instrument of what note length that instrument is currently playing.

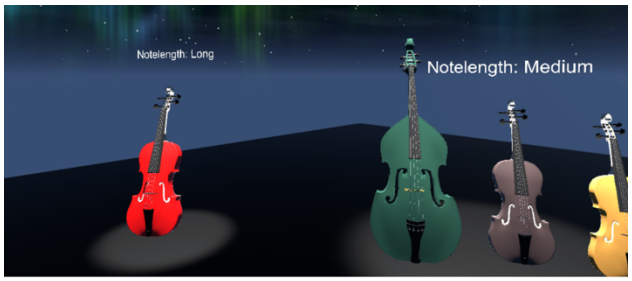

Figure 3: The note length representation inside the scene.

To perform these actions in the game we have the user point at an instrument, press and hold button 1 on the left controller and move it to the left or right to change the note length of a single instrument. To acquire the same effect on all instruments simultaneously, the user presses the touchpad on the left controller and moves the it to the left or right. To shorten the notes the user moves the controller to the right since the distance between the left and right hand becomes shorter. To lengthen the notes, the user moves the controller to the right to shorten the distance between the two controllers.

Midi Visualization

Most people do not know how to read scores and therefore some kind of visualization is needed that would let the user know what instruments should be active at each section in the song. A potential situation would be that a user mutes all the instruments and wants to bring one instrument into play. If there is no visualization, the user has no way of knowing which instruments are active at that time and must guess until an active instrument is found.

One possible solution is capturing images of the MIDI files and displaying them above each instrument. However, the length of the song would require us to capture multiple images for each instrument. We would have to account for the tempo of the song which is constantly changing and therefore this solution is too tedious and time consuming to be considered.

Wwise does not offer any functions that are called when a MIDI file is actively playing, neither do they offer variables that are set according to the music's state. What we did was to manually find the time each instrument starts playing and the time the corresponding music segments ends, meaning that each instrument can have more than one segment and is not necessarily playing throughout the whole song.

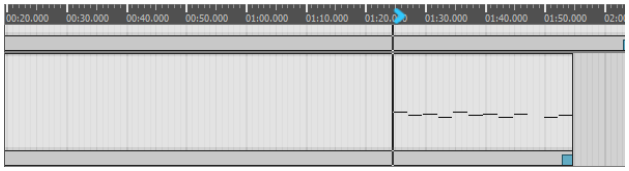

Figure 4: MIDI file represented in Wwise

For this solution, each instrument has a component that keeps track of the times in a list of floating point numbers. The tempo is constantly changing and therefore the times must be scaled accordingly. To achieve this, we check whether the tempo has changed and if so, change factor of the tempo is calculated and is multiplied with every time

variable that the list contains. Now that the seconds are correctly calculated, we subtract the delta time from each variable (delta time is the time it took to render the last frame and we do this calculation every frame). This way the float numbers would accurately approach zero for when the instrument starts/stops playing.

The times are stored in lists, one for start times and one for end times. At every frame, the first variable in the start times list is checked whether it has reached 0. If it has, a spotlight is activated above the instrument in question and the start time variable is removed from the list. This puts the next start time at the first index of the list, so the first variable is the only one checked even though the instrument has many music segments. We do the same thing for the end time lists, but it is only checked if the instrument is active.

User Testing

The goal of our study was to figure out in what way it would be best to interpret the movements of a conductor in virtual reality. The hand gestures should be intuitive for novice users and simple enough for the user to do two or more simultaneously.

Method

During the user tests we wanted to test five different core features we had developed for the game:

- Manipulate individual instrument volume
- Manipulate the volume of all instruments simultaneously
- Manipulate individual instrument note length
- Manipulate the note lengths of all instruments simultaneously
- Manipulate song tempo (BPM)

We asked 10 people with no experience in conducting an orchestra or any musical group for that matter to participate separately in the test. The test was split up in to three sessions.

The first session was started by teaching each tester how to operate the system. At this point in the test we did not want the tester to be using both hands, so we had them play around in the system and ask them to do all the tasks that require only the left hand (table 1).

After this first part of the test we interviewed the participants and asked them to answer, on a scale 1-5 (1 being lowest), how intuitive it was to perform each required action on two scales:

- Button
- Gesture

In the second session we asked them not to do any of the tasks they had done in the previous session but only the right hand gestures (table 1).

After this second part of the test we interviewed the participants again and asked them to answer, on the scale of 1-5, how intuitive it was to perform the required action on two scales:

- Button
- Gesture

In the third session we asked participants to combine the previous two sessions, use both hands at the same time and simultaneously keep the tempo while playing around with the other operations (all operations in table 1).

After this third and last part of the test we interviewed the participants once again and asked them to answer if it was required more effort to perform the required actions while simultaneously keeping tempo and to rate how intuitive each operation felt on two scales:

- Button
- Gesture

During each session we asked each participant to rate how engaging each operation was on a scale of 1-5 and at the end of the session how much they liked playing on a scale of 1-5. We also asked the participants to think out loud while testing the game.

Operation	Hand	Button	Gesture
Change single instrument volume	Left	Trigger (button 7 on figure 1)	Raise/lower hand
Change all instruments volume	Left	Grip (button 8 on figure 1)	Raise/lower hand
Change individual instrument note length	Left	Button 2 (button 1 on figure 1)	Move hand left to increase Move hand right to decrease
Change all instruments note length	Left	Touchpad (button 2 on figure 1)	Move hand left to increase Move hand right to decrease
Manipulate tempo	Right	Grip (button 8 on figure 1)	Swing controller from back and forth from left to right

Table 1: All the available operations in the game

Results

The outcome of the test was a list of ratings on the scale of 1-5 of how difficult and engaging each operation was for the participants and how much they enjoyed playing each session of the test. We took the mean from the ratings gathered for both buttons and gestures and added it together to get a scale of 1-10 and used it as the final test results. The results from the sessions are as follows:

Operation	Button rating (1-5)	Gesture rating (1-5)	Combined rating
Change single instrument volume	5	4,4	9,5
Change all instruments volume	5	4,5	9,5
Change individual instrument note lengths	3,1	3,7	6,8
Change all instruments note lengths	3,9	3,7	7,6

Table 2: Test results from session 1

Operation	Button rating (1-5)	Gesture rating (1-5)	Combined rating
Manipulate tempo	5	3,6	8,6

Table 3: Test results from session 2

Operation	Button rating (1-5)	Gesture rating (1-5)	Combined rating
Change single instrument volume	4,6	3,9	8,5
Change all instruments volume	4,5	3,9	8,4
Change individual instrument note lengths	2,9	3,5	6,4
Change all instruments note lengths	3,5	3,5	7
Manipulate tempo	4,5	3	7,5

Table 4: Test results from session 3

Session	Level of enjoyment (1-5)
1	2,6
2	1,9
3	3,2

Table 5: Results from how much users enjoyed playing

Discussion

One way to interpret the results from these tests is to say that as the first two sessions are combined in session 3, the game gets less intuitive. At the same time, the game is more fun to play. The rating for intuitiveness on the operations performed in session 1 is lower by an average of 0,775 on each operation as they are played again in session 3 simultaneously with the tempo manipulation operation which got a rating 1,1 lower in session 3 than in session 1. On the other hand, the rating for how fun it was to play doubles in between session 1 and 3. From that we can gather, the participants in the test thought it weighed more that the game was fun at the expense of being a little less intuitive.

Future Work

The results show that a novice user can conduct an orchestra to some extent with modern virtual reality technology. However, capturing the emotion of the conductor is a big part of conducting and won't be possible without significant advances in virtual reality technology.

Conclusion

A novice user can conduct an orchestra in virtual reality but only to a certain extent. All of the core features were implemented and tested successfully. However, conducting a real orchestra is vastly different from our application since we are missing all the factors that humans possess but machines don't.

Acknowledgements

We appreciate the help we got from the people that tested the game for us, both with the testing and interviews.

References

Malaika, Y. 2015. Interaction Design in VR: The Rules Have Changed (Again). GDC Europe Online: https://www.gdcvault.com/play/1022810/Interaction-Design-in-VR-The/.

Butorac, D. 2015. The Language of Conducting. https://www.youtube.com/watch?v=xcR1-WhjZys/.

Schwarz, G. 2014. Gerald Schwarz Gives a Conducting Lesson. https://www.youtube.com/watch?v=vivT5NaT8Kg/.

Zander, B. 2017. Interpretation Class: Beethoven – Symphony No. 5, Mvt. 1. https://www.youtube.com/watch?v=JbOGstahWKI/.

Wolf, K.; Naumann, A.; Rohs, M.; and Müller, J. 2011. A Taxonomy of Microinteractions: Defining Microgestrues Based on Ergonomic and Scenario-Dependent Requirements. In *Human-Computer Interaction – INTERACT 2011*, 559-575.

Robertson, D. 2010. The Art of Conducting? https://www.youtube.com/watch?v=xKak7V6al1I/.