3DJ: Interaction Techniques for the Manipulation of Audio Within 3D Environments

Elie Diaz
Georgia Institute of Technology
Atlanta, GA
ediaz30@gatech.edu

Russell Strauss
Georgia Institute of Technology
Atlanta, GA
jstrauss6@gatech.edu

ABSTRACT

October 14, 2020. With an increasingly complex industry revolving around audio mixing and editing, the need for useful interactions that accomplish these tasks in a 3D virtual space continues to grow. Our project explores methods for sound and music interaction. This involves the development of different ways of visualizing the audio as well as seeing which actions feel most natural to users in this 3D space. While work has been done on sound interfaces, much of the work stays in the 2D space for digital audio workstations. However, there are several tools that exist for users to work with instruments in a VR space. Our project develops a tool set that allows users to manipulate and create audio in a 3D setting, combining some of the work done for digital audio workstations in 2D with the richness of a 3D space seen in VR. The metaphor we explore is a DJ-style workstation, such that the user can observe and interact with several different tools that change the sound being played in real time.

INTRODUCTION

We built this project for exploring novel methods of sound and music interaction in 3D space. We have built a DJ-style workstation to provide the user a meaningful context for their actions, different functions for manipulation, and a designated space to interact. Immersive musical instruments give users a high degree of control with gestural interaction to perform, compose, or improvise music in real time. [9] We have also enabled ways to select and manipulate specific audio attributes such as level adjustments, effect filters, changing playback rate, and playback control. These basic uses have application in a wide variety of tools from simple audio editing to live music creation. Metaphors are commonly used in UI as powerful tools to enable quick understanding and more effective usability [2], and thus we chose to simulate a virtual digital audio workstation. We have bridged the gap between the actions a user can perform and their natural human interaction. For example, using the distance between the user's hands to adjust values creates a more natural interaction than selecting a volume modifier from a 2D menu.

CONTROLS

- Adjust playback position: right hand–use a ray to point to desired position in waveform and click or drag trigger
- Adjust playback rate: left hand–position hand on turntable and pull trigger to rotate clockwise or counterclockwise
- Select record: right hand– point to album cover and pull trigger, then place record upon turntable with left hand to begin playback
- Select an effect toggle: right hand–position 3D cursor of right hand upon toggle block and pull trigger
- Adjust level-fader: right hand: position 3D cursor of right hand upon level adjustment block and hold down trigger while moving hands closer or farther from each other to adjust levels

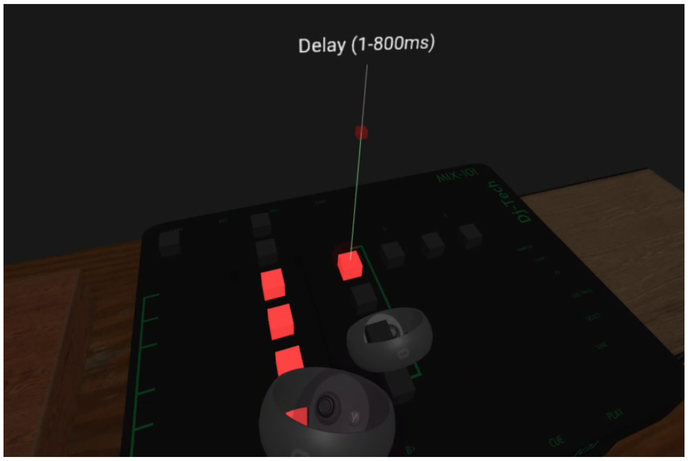

Figure 1. Adjusting values in 3D using the balloon technique.

INTERACTIONS

At its basics, the visualized waveform allows you to identify the start and end of the track for alignment, the repetitive rhythms as volume spikes for beat recognition, and the overall structure of the composition to identify sections of the audio such as a long, quiet intro or using peak and valley locations to identify the positions of song transitions or points of interest in a piece of audio. Although outside the scope of the project, you could also imagine mapping audio attributes to other multidimensional visuals such as visualizing a key or the pitches of notes in contrast to the classical 2-dimensional aspect of sound wave sampling in the time-vs-magnitude form.

The 3D nature of the interface allows for more effective affordances using hand gestures alternative to dragging a slider with a mouse. Handheld controls contribute high-precision advantages over their 2-dimensional counterparts. Visualizations of waveforms provide insight to the user on details of the selected audio and the context in which it will interact with the main audio output being composed. For our sound visualization, the audio is parameterized into samples that are mapped to visual elements. The visualization of sound allows us to take advantage of the strong pattern recognition abilities of the human visual system to identify patterns and structure of audio signals. [3]

The organization of our interface provides relevant information to the user. The level-adjusters connected to the track couple the actions of that interactor to its specified effect and their subcomponents. The colors of hover state (blue) vs. active state (red) allows the user to quickly identify the behavior of the system without having to think–cognitive load is extremely important in real-time interactions such as music manipulation. The effects can be toggled, but their subcomponents also modified–the red active state shows the user the toggle-state of the effect, while the blue shows the user which level-adjustment block will be affected based on his current hand location. It is important that the user knows exactly which effector he is adjusting since clicking the wrong effect adjuster at the climax timing of a song beat will have dramatic effects against the sound output the user aims to achieve.

The concept of sound objects using 3D shapes is a straightforward idea that allows interaction to become an easy task. It also enables the user to expand musical thoughts in new ways of composition and performance. [5, 9] The user may adjust the playback rate of the audio using the turntable object interactor. Similar to a physical turntable, the user can rotate the vinyl record in place to speed it up or slow it down. The rotational speed of the record in combination with the audio output communicate the state of the playback rate manipulation. Similarly, if the user wishes to navigate to a specific point in the song, he can point to the exact location in the waveform object and pull the trigger or scrub back and forth to navigate to a precise audio playback location.

Existing sound editing tools are designed specifically for audio engineers. Sound generation tools from game libraries and computer animation software allow computer graphics professionals to integrate sounds into their work, but these tools provide an automatic process to add sound and do not consider human factors or allow for inspiration through the tools. [4, 6, 8] The existing tools do not address 3D interaction techniques or human-factor input for audio. It is difficult to learn to create and control high quality sounds using digital synthesizers. The interfaces have high learning curves and thus musicians often rely on default factory settings. Musicians have refined motor skills that can be more heavily utilized in the context of a 3D gestural system by continuous movements that reduce cognitive load. [10] When surveying users on 3D virtual music spaces, researchers noted interactions for playable instruments as a common desire among users. [1] This high learning curve can act as a stifling factor for creativity. For example, looking at a mathematical algorithm for an audio manipulation effect may not provide any insight. However, when the user can dynamically adjust these level in combination with other manipulators, it provides extremely rapid feedback on the permutations of output contained within those algorithms. This allows the user to improvise based upon that feedback loop and quickly identify effective relationships between manipulators to create interesting and inspiring manipulations of musical pieces.

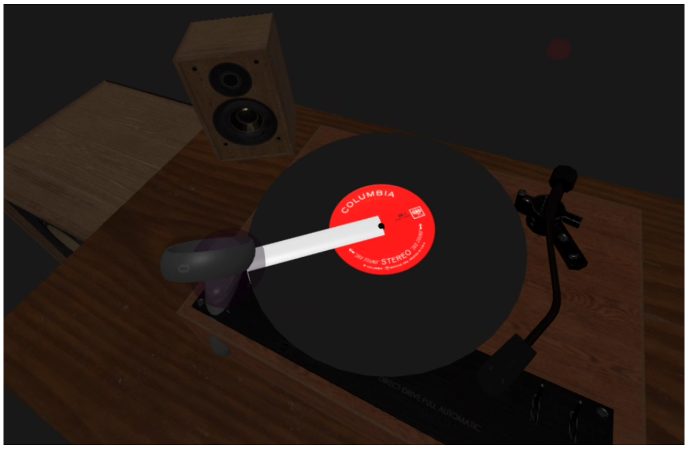

Figure 2. Adjusting playback rate using rotational interactions.

RESULTS AND CONCLUSIONS

The resulting experience provides a system in which users are familiar with certain elements and can utilize their existing knowledge to guide their interpretation of the system. For example, many know that turntables progress by rotating a disc made of vinyl. This will inform the user's interaction decisions as he begins to learn the interactions. A visual feedback indicator (Figure 2) and the user's previous knowledge work together to communicate the behavior of the interactor.

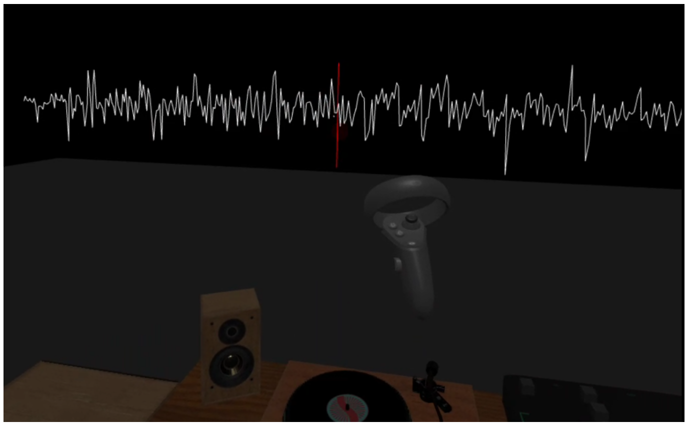

Figure 3. Navigating audio with ray-casting by allowing the user to point to a specific location using hand gestures.

In a similar way, we have referenced standards developed by existing tools in audio manipulation. For example, we display our audio visualization in a classical sample-based audio waveform–users already understand existing standards and bring expectations for how the system might work before ever interacting. We wish to utilize these expectations instead of swimming against them upstream. As such, the waveform and progress navigation behaves in ways typical to prior audio manipulation tools, however the user is interacting with more advanced inputs. Thus he can point to a specific point to navigate to the location precisely. A picking indicator dot shows exactly where the click will occur before committing. The natural manner in which he interacts with the waveform with his gestures gives him additional accuracy to make quick and accurate selections–compared to physically moving a mouse across a table to coordinate a selection on a separate screen. Fitts's law shows that users are limited by the speed of their physical response and reaction times when they are required to move their cursors long distances quickly to a small specific location. It is important to compensate for this factor to provide a system that will allow live performance interactions in sync with the tempo of the selected music. The gestural interactions of our system decrease this greatly because the user's actions are mapped 1-to-1 to his natural body movement in the physical world.

With all of the interactions in play, our question becomes: "How effective is our application? Can it lead to novel output on behalf of the users?" The user's interactions have real-time effects on the music while using our application. This allows users to continue experimenting with different effects to manipulate the audio in ways that achieve a desired result. This experimentation adds to the quality and feel of our application: for example, Steve Swink identifies that virtual sensation "requires an exploration of some of the ways people perceive things, including measures for frame rate, response time and other conditions necessary for game feel to occur." The three building blocks of game feel are real-time control, simulated space, and polish-quality with carefully selected finishing touches of the experience. [7] These elements are important for an application such as ours where the experience of the creation is itself is an important parameter to the application. Using these definitions as a litmus test, we judge our prototype effective: the user is immersed in a virtual creative space which leads to new experiences and novel music creation while providing a fun and effective 3D virtual system with which to interact with natural gestures.

IMPROVEMENTS AND EXTENSIONS

From testing with some new users and our own experiences using the application, there are several ways in which the application can be extended to further explore audio manipulation in 3D. For example:

- Add improved playback control with designated button for playing, stopping, and restarting playback.
- Scale the functionality for two records which can be manipulated and mixed in relation to each other. The program was designed in an object-oriented fashion to enable the eventual possibility of controlling two records simultaneously, each Record object containing all information needed for the audio playback and manipulation: the audio file, state, playback rate, chained effects with their various input values, a sound waveform, and anything else needed by the audio track.
- Adding programmatic BPM analysis and beat recognition for synchronizing rhythms and audio tracks relative to each other for mixing
- Provide audio clip highlighting for trimming or selecting subsections of audio
- Optimize performance: performance was outside the scope of the project, so long that it did not affect the user experience or interface. To load more than 2 songs simultaneously, some work will need to be done to optimize the audio loading. The ability to add a user uploaded song is another desired extension that will require thought since the workflow to add media from a VR-device could require much effort from the user. We could create a desktop uploader that allows you to add media from your computer to then access it in runtime from the VR device.
- Add some details to the existing UI such that actions feel more natural to a user that has never done them before. This includes such details as a slight rumble in the controller when the user's pointer hovers over key items, as well as visual indicators that the disks can be touched and rotated before the user does so.
- Add a wider variety of UI mechanisms: our 3D sliders and rotational interactors are good for single-value attributes, but I would like to explore 3DUI's that encompass other ranges of inputs. For example, we may add a list of songs or users may upload songs–a scenario in which you would need a method for selecting items from a list. It may also add value to allow the user to pick and choose which manipulators (effects) that he wishes to attach to his customized board and set a mixer layout unique to his needs.
- Add the functionality to save and export the audio from a session. This can allow users to share the mixed audio outputs with others.
- Add ability to export states of the effects and playback settings used in a previous session.

REFERENCES

[1] Leena Arhippainen, Minna Pakanen, and Seamus Hickey. 2012. Designing 3D Virtual Music Club Spaces by Utilizing Mixed UX Methods: From Sketches to Self-Expression Method. In *Proceeding of the 16th International Academic MindTrek Conference (MindTrek '12)*. Association for Computing Machinery, New York, NY, USA, 178–184. DOI: http://dx.doi.org/10.1145/2393132.2393167

[2] Doug A. Bowman, Ernst Kruijff, Joseph J. Laviola Jr, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional.

[3] Matthew Cooper, Jonathan Foote, Elias Pampalk, and George Tzanetakis. 2006. Visualization in Audio-Based Music Information Retrieval. *Computer Music Journal* 30, 2 (2006), 42–62. http://www.jstor.org/stable/3682003

[4] Holger Hennig, Ragnar Fleischmann, Anneke Fredebohm, York Hagmayer, Jan Nagler, Annette Witt, Fabian J Theis, and Theo Geisel. 2011. The Nature and Perception of Fluctuations in Human Musical Rhythms (Nature of Fluctuations in Human Musical Rhythms). *PLoS ONE* 6, 10 (2011), e26457.

[5] A. Pirhonen, S. Brewster, and C. Holguin. 2002. Gestural and audio metaphors as a means of control for mobile devices. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 4. 291–298.

[6] Christian Sauer, Thomas Roth-Berghofer, Nino Auricchio, and Sam Proctor. 2013. Recommending Audio Mixing Workflows. In *Case-Based Reasoning Research and Development*, Sarah Jane Delany and Santiago Ontañón (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 299–313.

[7] Steve Swink. 2008. *Game Feel: A Game Designer's Guide to Virtual Sensation*.

[8] M.K.-P Tong and Kam-Wah Wong. 2006. Sound Editing Interface For Computer Animators. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 5. IEEE, 3629–3634.

[9] L. Valbom and A. Marcos. 2007. An Immersive Musical Instrument Prototype. *IEEE Computer Graphics and Applications* 27, 4 (July 2007), 14–19. DOI: http://dx.doi.org/10.1109/MCG.2007.76

[10] Roel Vertegaal and Ernst Bonis. 1994. ISEE: An Intuitive Sound Editing Environment. *Computer Music Journal* 18, 2 (1994), 21–29.