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

April 19, 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. [8] 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

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, 8] 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. If the user wishes to navigate to a specific point in the song, he can point to the exact location in the waveform and pull the trigger or scrub back and forth to navigate to a precise audio playback location.

Tong notes that 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, 7] 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. [9] 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 manipulaions of musical pieces.

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] 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.

[8] 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

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