



**HAL**  
open science

## Comparison of Spatial and Temporal Interaction Techniques for 3D Audio Trajectory Authoring

Justin Dan Mathew, Stéphane Huot, Brian F. G. Katz

► **To cite this version:**

Justin Dan Mathew, Stéphane Huot, Brian F. G. Katz. Comparison of Spatial and Temporal Interaction Techniques for 3D Audio Trajectory Authoring. *Journal on Multimodal User Interfaces*, 2020, 14, pp.83-100. 10.1007/s12193-019-00314-x . hal-02375493

**HAL Id: hal-02375493**

**<https://inria.hal.science/hal-02375493v1>**

Submitted on 22 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comparison of Spatial and Temporal Interaction Techniques for 3D Audio Trajectory Authoring

Justin D. Mathew · Stéphane Huot · Brian F.G. Katz

Received: October 2018 / Accepted: November 2019

**Abstract** With the popularity of immersive media, developing usable tools for content development is important for the production process. In the context of 3D audio production, user interfaces for authoring and editing 3D audio trajectories enable content developers, composers, practitioners, and recording and mixing engineers to define how audio sources travel in time. However, common interaction techniques in 3D audio production tools can make the workflow of this task tedious and difficult to accomplish. This study investigates this problem by classifying the atomic tasks (spatially and temporally) of a general composite task of authoring 3D audio trajectories and then evaluating different interaction techniques across these tasks. Common graphical user interfaces were compared with input devices having varying degrees-of-freedom for spatial atomic tasks in order to investigate the effect of direct manipulation and integrality of interaction techniques. Continuous and discrete interaction techniques were compared for temporal tasks in order to investigate the effect of

direct manipulation. Results suggest that interaction techniques with high degrees of integrality and direct manipulation reduce task completion time compared to standard GUI techniques. The design of temporal tasks can create a visual bias, and discrete-time controls can be a suitable method for traversing a small number of control points. These results and further observations provide directions on the study of interaction technique design for 3D audio tools, which in turn should improve workflows of 3D audio content creation.

**Keywords** 3DUI · 3D Audio · Human-Computer Interaction

## 1 Introduction

The advent of media technologies has provided users with the capability to experience a vast amount of three-dimensional content. Recent trends in the development of immersive media technologies have been increasingly popular with consumers. Currently, immersive technologies improve spatial experience through visual (*e.g.*, large 2D visual displays, 3D glasses, virtual reality headsets, etc.) and auditory displays (*e.g.*, binaural headphones, 3D speaker configuration, etc.). If content producers are required to create an experience that is time-based, such as 3D animation or 3D audio content, they are often required to define how 3D objects move in time. This requires the definition of spatial and temporal properties of the object's movement in their projects. A common example is the authoring of 3D trajectories, where one defines the 3D path and movement in time of an object. Standard 3D content production systems

---

Justin D. Mathew  
Inria Lille-Nord Europe, LIMSI-CNRS, Univ. Paris-Saclay  
Fort Lauderdale, FL 33304 U.S.A  
E-mail: justin.dan.mathew@gmail.com

Stéphane Huot  
Inria, Univ. Lille, UMR 9189 - CRISTAL  
40 avenue Halley  
59650 Villeneuve d'Ascq, France  
E-mail: Stephane.Huot@inria.fr

Brian FG Katz  
Sorbonne Université, CNRS, Institut d'Alembert, UMR 7190,  
Lutherie - Acoustique - Musique  
4 place Jussieu 75005 Paris  
E-mail: brian.katz@sorbonne-universite.fr

---

<sup>0</sup> A portion of this work was included in the first author's PhD thesis (Mathew, 2017)

are often based on interaction techniques and graphical user interface (GUI) elements that have originally been designed for common mouse and keyboard devices, usually augmented with drawing and editing automation lines along a timeline or sequencer. This can be seen in popular Digital Audio Workstation (DAW) audio production software, 3D animation software, and video game development software, where one can define the change of a single parameter along a timeline.

Users can employ these techniques to define how an object’s position will change along a specific axis by defining and editing key frames or control points along the sequencer with their mouse cursor. These control points define the positional values of the trajectory path over time, and users can edit these values along the sequencer’s  $y$ -axis and when they occurs along the sequencer’s  $x$ -axis. However, current users of these common tools have described this interaction paradigm as limiting and tedious when authoring 3D spatial parameters (Mathew et al, 2017). This forces one to define the change of individual spatial parameters (*e.g.*,  $x$ -position vs. time,  $y$ -position vs. time,  $z$ -position vs. time, etc.) separately rather than defining the change of multiple spatial parameters simultaneously (*e.g.*,  $\{x, y, z\}$ -position vs. time). This is likely due to the lack of *integrality* and *direct manipulation* of the interaction technique available in these tools.

*Integrality* describes the comparison between the number of parameters required to manipulate within a task (structure of the task) and the degrees of freedom (DoF) of each input method available (structure of the input method) to conduct the task (Jacob et al, 1994). For example, a device, such as a 2DoF touchscreen, is highly integral when the user is tasked to use one finger to point, select, or drag objects across the screen.

*Direct manipulation* describes how closely the range of values generated by the input method is mapped to the parameters of the object of interest to manipulate (Aceituno, 2015; Shneiderman, 1997; Damaske, 1969). Using the previous example, a touchscreen for touch interactions on the screen is considered an interaction technique with a high degree of direct manipulation. Using an external input device, such as a mouse, to control a cursor on the screen that is then used to point, select, or drag objects of interest would be considered a technique with less direct manipulation.

It has been suggested that increasing both the integrality and direct manipulation in the design of interaction techniques improves user performance in conducting spatial tasks (Jacob et al, 1994; Martinet et al, 2012, 2010). Even though input devices with a high number of DoF have been proposed and studied for 3D manipulation tasks, they have not been studied thoroughly in

the context of authoring spatial parameters for immersive media content. Our hypothesis is that employing interaction techniques with high degrees of integrality and direct manipulation would also improve *usability*, which is described as the ease of use and ability to learn a tool (Nielsen, 1993), in conducting tasks related to 3D audio trajectory authoring.

This work investigates the influence of different user interface designs on the usability of authoring 3D trajectories in the context of 3D audio production. We first present a classification of the atomic tasks that relate to the general composite task of authoring 3D audio trajectories. Secondly, we report the results of a user study that compared different interaction techniques with varying designs and integrality across these atomic tasks. In this study, participants were placed in a setting that simulated how practitioners produce 3D audio content. They were asked to use five spatial and three temporal interaction techniques to conduct three tasks. Spatial interaction techniques consisted of two standard GUI techniques that used a trackpad to control the mouse cursor and three direct manipulation techniques with varying DoF. Temporal interaction techniques consisted of two standard GUI techniques with different mappings where participants used the mouse cursor to interact with a timeline/sequencer and a direct manipulation technique that discretely navigated the timeline (swipe gesture on the trackpad).

Results identified that specific design characteristics of the tasks and interaction techniques reduced the average completion time. As expected, participants performed faster and preferred interaction techniques that employ direct manipulation and have high integrality. In addition, the structure and design of the spatial tasks had an effect on completion time. Results from the temporal tasks highlighted a visual bias, which allowed participants to rely on their visual rather than auditory reference. However, it did show potential use cases for a discrete interaction technique. In addition to these results, participant comments and observations provided explanations on their performance of tasks and preferences. With these results in mind, we present a preliminary discussion on how further studies could combine atomic spatial and temporal interaction techniques for the purpose of studying their potential use in conducting composite tasks. Based on these results, we propose future directions on the study of tasks and user interface designs for authoring 3D audio trajectories and other types of time-based 3D digital content.

## 2 Related Work

In the context of 3D audio production, standard tools allow interactions with spatial and temporal parameters through mouse-based drawing and editing of automation lines (Mathew et al, 2017). This commonly used interaction technique was analyzed by Metatla et al (2016), who classified atomic tasks related to interacting with automation lines to investigate the use of haptic interfaces for the visually impaired. In that study, the authors defined specific tasks such as locating an existing or adding a new control point, estimating an existing or new control point's target position, and switching between control modes to place the control point at the target position. In these tasks, the standard mouse-based interaction techniques separate the 3D positional control, but there has been interest in advancing this for more integral interaction techniques. A popular method in 3D graphics is to use the mouse and ray-casting techniques (Van Emmerik, 1990; Schumacher and Bresson, 2010; Bresson and Schumacher, 2011), but this is not widely used in consumer based 3D audio production tools. Available 3D audio software tools, such as 3DC Lib in OpenMusic (Bresson and Schumacher, 2011), Spat library for Max (Carpentier, 2015), and Spatial Audio Workstation <sup>1</sup>, provide multiple 2D planes that represent the same 3D space, allowing users to compose and edit 3D audio trajectories.

There has been interest in using other input devices to improve user interactions. In the same OpenMusic programming environment, Garcia et al (2015) used interviews and participatory design to contribute to the development of Open Sound Control (OSC) functionality for a communication protocol that allows various external input devices to control positional parameters, such as the GameTrak controller. Favory et al (2015) then extended this work by investigating the use of a multi-touch tablet. In both of these user studies, observations and participant feedback were collected to propose directions for improving the interaction and visualization. Melchior et al (2013) compared a desktop mouse and a 6-DoF input device with haptic feedback for the task of creating trajectories by defining position and velocity either separately or simultaneously. Using heuristic ratings, participants rated the mouse-based interaction technique as having better predictability and being easier to manage, but the majority still preferred the 6-DoF input device overall. Even though studies have investigated interaction techniques for authoring and editing 3D audio and visual trajectories, results have focused more on qualitative evalu-

ations than quantitative results. Outside of audio production, 3D input devices have mostly been compared to each other through their usability for completing a 3D manipulation task (Zhai and Milgram, 1998; Jacob et al, 1994; Martinet et al, 2012, 2010; Veit et al, 2011). Authoring techniques of 3D paths have mostly been studied through mouse-based interfaces using sequential techniques that define trajectory paths through gestures (Zeleznik et al, 2007), constraints (De Araujo et al, 2012), and interpretations of user-drawn strokes or constraints (Sachs et al, 1991).

In the context of interacting with the temporal dimension within audio and music production, time has been commonly represented along a timeline or sequencer using a mouse and cursor for interaction. Outside of audio production, there have been more interests in visually representing time in different forms rather than interaction techniques. Circular forms, like a clock, use rotation and angles to represent time and events, which can also be seen in spiral timelines. Applications, like the SpiraClock (Dragicevic and Huot, 2002) and Conversation Clock (Bergstrom and Karahalios, 2007), explored event notifications and audio conversations respectively. Another method represented used frames of images in a linear sequence to show the change of spatial properties over time (Garcia et al, 2015), but this technique does not explicitly state the duration within the image. Bresson and Schumacher (2011) proposed a dynamic parametric spacing between 3D points to define how a 3D audio trajectory will be temporally sampled, but most have simply utilized the common straight timeline. In more advanced media systems, such as Unity 3D <sup>2</sup>, AutoDesk 3ds Max <sup>3</sup>, or Iannix software (Jacquemin et al, 2012), an interactive timeline coupled with another graphical view shows how higher dimensional media (*e.g.*, video, 2D or 3D animation) change over time. In the majority of these instances, time is defined along a single dimension that moves forward in one direction and backward in the opposite direction, and temporal parameters can be manipulated along that dimension. The temporal parameter is typically controlled along a single dimension using mouse-based interaction techniques, allowing users to define when spatial parameters change. However, advances in interaction techniques have only focused on the visual representations and have not explored techniques for navigating along a timeline.

<sup>2</sup> <http://www.unity3d.com/>

<sup>3</sup> <https://www.autodesk.com/products/3ds-max/overview/>

<sup>1</sup> <https://facebook360.fb.com/spatial-workstation/>

### 3 Methodology

Based on previous works (Jacob et al, 1994; Martinet et al, 2012, 2010), we hypothesize that increasing the direct manipulation and DoF of the input method in the interaction technique is likely to reduce the completion time when conducting 3D manipulation tasks. This study examines the hypothesis that increasing the direct manipulation and integrality of the interaction techniques would also improve task completion time, be easier to learn, and be more preferred in the context of 3D audio production. In order to test this, we first identified that a description of the general composite task and its decomposed atomic tasks are required. This was similarly done in the context of authoring and editing automation lines in Digital Audio Workstations (Metatla et al, 2016). We adopted this strategy by describing the composite task of authoring and editing a 3D audio trajectory as requiring users to interact with spatial and temporal properties by switching between defining new and editing existing control points that describe the trajectory. This can be decomposed into the atomic tasks of placing a new control point at a target location by (1) *adding* a new point in a blank project or extending an existing trajectory or (2) *inserting* a new point into an existing trajectory. Editing 3D audio trajectories often requires first *selecting* an existing control point before manipulating it.

Using this classification of atomic tasks, we compared different interaction techniques by evaluating them across each task in a user study. In order to investigate direct manipulation and integrality in this context, interaction techniques were chosen to be indirect manipulation techniques commonly found in consumer 3D audio software tools and input devices that can directly manipulate and have varying DoF to test from fully separate to fully integral interaction techniques. An initial pilot study with members of the research team investigated the possibility of combining spatial and temporal tasks into one task, but it became quickly obvious after a couple of participants that it was difficult to compare the interaction techniques as participants were required to switch between two interaction paradigms. For this reason, the study was split into two sessions that asked participants to conduct the spatial and temporal atomic tasks independently. Separating these tasks reduced the complexity and provided a more controlled methodology to compare the design characteristics of the interaction techniques.

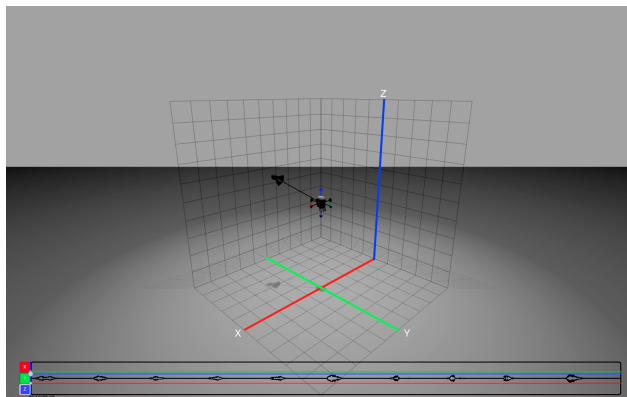


Fig. 1: GUI used to test each interaction technique.

#### 3.1 Experimental Setup

Participants were situated at a desk with a Macbook Pro 13-inch laptop and headphones that simulated how practitioners would conduct these tasks in a common production scenario. In this setting, participants were presented with a 3D graphic scene on the laptop screen and a 3D audio mix that consisted of a female voice counting from zero to nine (indicating the time in seconds). If the female voice said the number four, it would occur at 4 s into the audio file. The voice audio file was binaurally rendered and then mixed with binaural recordings of a drum ensemble and environmental noise in an effort to increase the sense of immersion. Participants were presented with a 1 s loop (audio playback window) of the 3D audio mix wherever the center of the playback window (temporal cursor) was located ( $\pm 0.5$  s of the position of the temporal cursor).

On the 13-inch laptop display, the 3D virtual scene comprised a virtual  $2 \times 2$  m cube to define the range of positional parameters and 3D model of a head to indicate the origin of the scene and listener location (see Figure 1). The cube provided grid lines that indicated distance intervals across each dimension, where each grid value was equal to a virtual distance of 0.2 m. In total, the range of values for each dimension was from  $-1$  to  $+1$  virtual meters from the center of the scene. Lighting of the 3D virtual scene was placed right above the center of the scene, creating shadows of the 3D objects presented in the scene. Participants were allowed to rotate the camera viewport around the 3D graphics scene for a different perspective. The range of camera rotation was constrained between a bird's eye view of the transverse plane (x & y axes), frontal plane from behind the head (x & z axes), and the median plane from the left and right of the head (y & z axes). This was constrained in order to prevent participants

from losing their frame of reference in the 3D virtual scene.

Below the 3D virtual scene, an interactive timeline was provided that included a visual representation of the audio waveform of the female voice and temporal properties of the 3D audio trajectories. Within this, a light grey set of control points with colored borderlines (red for  $x$ , green for  $y$ , blue for  $z$ ) represented the position of the 3D cursor within the timeline and reacted in real-time when changing the position of the 3D cursor.

Instructions on which task they should conduct was presented above the virtual scene. For interaction techniques that used only the trackpad and keyboard for input methods, participants were just provided with the laptop and a pair of headphones. For the interaction techniques that used external input devices, the input method was placed close to the trackpad and keyboard in order to quickly access the arrow keys for camera rotation (see Figures 5 and 6). Camera movements only affected the rendering of the visual 3D virtual scene and did not affect the rendered 3D audio playback. Participants were instructed that the 3D model head in the scene represents a stationary listener, and the 3D audio playback was rendered with respect to that listener.

### 3.2 Tasks

Tasks were designed to evaluate participant completion time when controlling a spatial and temporal cursor in the 3D virtual scene and GUI timeline. Participants were instructed to directly control the position of a cursor in order to place it at the target location for adding, inserting, or selecting control points, providing consistency between the tasks.

In the spatial tasks, any spatial manipulation would result in a change in the 3D position of the female's voice, but the temporal location within the audio mix remained constant. In the temporal tasks, participants were asked to navigate along the timeline and listen to the audio mix at different points in time to find the target time, but the spatial positioning remained consistent to the trajectory path given in the 3D graphic scene. Under these conditions, participant were asked to conduct three spatial tasks for five spatial interaction techniques and three temporal atomic tasks for three temporal interaction techniques detailed below.

#### *Add-Task:*

This task simulated the addition of a control point to an existing trajectory (see Figure 2).

- The goal of the spatial add-task was to position the 3D cursor to form a straight line with the distance between the 3D cursor  $C$  and control point  $B$

(distance  $BC$ ) to be equal to the distance between control point  $A$  and control  $B$  (distance  $AB$ ) while maintaining the order of *control point A (blue) - control point B (red) - 3D cursor C (grey)*.

- The goal of the temporal add-task was to place the temporal control points so that the time interval between control points  $A$  and  $B$  was equal to that between control point  $B$  and the temporal cursor  $C$ .

Control points  $A$  and  $B$  were placed at certain positions in time within the audio that represented moments when the female voice spoke the number associated with that time. Participants could identify the target location on the timeline by doubling the numbers said in the female voice at control point  $B$  and finding that location by listening and using the corresponding waveform presented in the timeline.

#### *Insert-Task:*

This task simulated editing a trajectory path by creating a new control point between existing control points. Participants were presented two control points ( $A$  and  $B$ ) with the 3D cursor  $C$  in between and connected to both points (see Figure 3).

- The goal of the spatial insert-task was to position the 3D cursor  $C$  halfway between the two control points to form a straight line where the distance between control point  $A$  and the 3D cursor (distance  $AC$ ) was equal to the distance between the 3D cursor and control point  $B$  (distance  $CB$ ) while keeping the order of *control point A (blue) - 3D cursor C (grey) - control point B (blue)*.
- The goal of the temporal insert-task was to position the temporal cursor  $C$  so that the time interval between control  $A$  and cursor was equal to that between control  $B$  and cursor  $C$ .

For the temporal Insert-Task, participants could identify the target location on the timeline by halving the numbers said in the female voice at control point  $B$  and finding that location by listening and using the corresponding waveform presented in the timeline.

#### *Select-Task:*

This task simulated the selection of a control point in an existing trajectory path (see Figure 4).

- The goal of the spatial select-task was to position the 3D cursor at the same location as the middle control point (control point  $B$ ).
- The goal of the temporal select-task was to select the middle control point (control point  $B$ ) on the timeline.

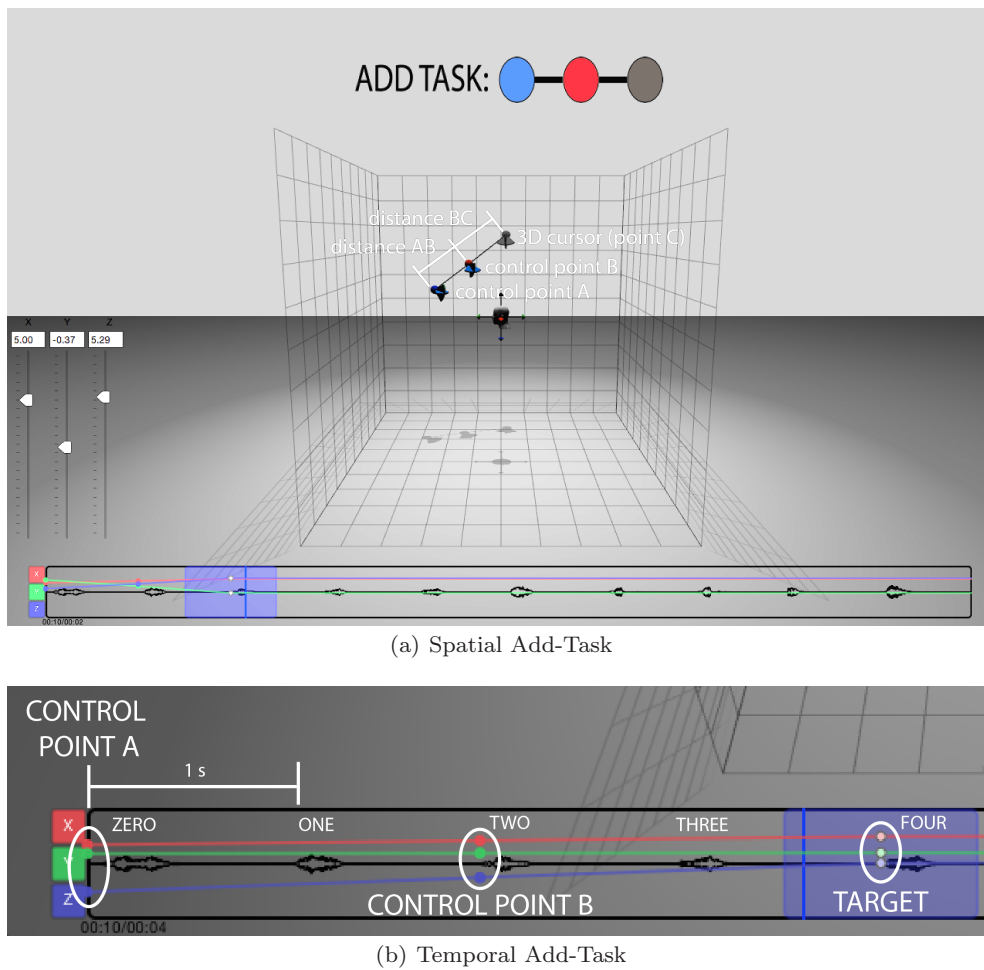


Fig. 2: Example of a spatial (a) and temporal (b) Add-Task.

In this task, participants were not required to use the audio content and could simply use the visual modality to select control point  $B$  in the timeline. Since these tasks were designed to use a cursor, manipulating an existing control point to a new location (a common editing task), was considered to be comparable to the Add-Task and Insert-Task.

The presented conditions and target position randomly changed per trial for all three tasks using a random generator function that was confined to ensure the given control points and target position were within the range of the cube. Within each interaction technique, participants were provided a method to validate their response using the down-click action of the trackpad, the enter key, or using a button that was on the external input device. If participants validated their response with the 3D and 2D cursor within the predefined error tolerance (0.3 m spatially within the virtual  $2 \times 2$  m cube and 0.3 s temporally), a green light flashed on the screen indicating the task was completed and partici-

pants moved to the next trial. If participants validated their response outside the error tolerance range, a red light flashed and participants were required to better position the 3D cursor and try again. The error tolerance, associated with the interaction techniques and visual feedback, was determined from a pilot study by trial and error, chosen based on pilot participant feedback and observations on their ability to complete the tasks.

### 3.3 Spatial Interaction Techniques

Participants conducted the spatial tasks using five different interaction techniques and were allowed to interact with them bi-manually. Two commonly used interaction techniques with indirect controls (GUI elements) were compared to three interaction techniques that employed direct manipulation and varied in DoF to vary the integrality of the interaction technique. We labeled these spatial interaction techniques as: *GUI Slid-*

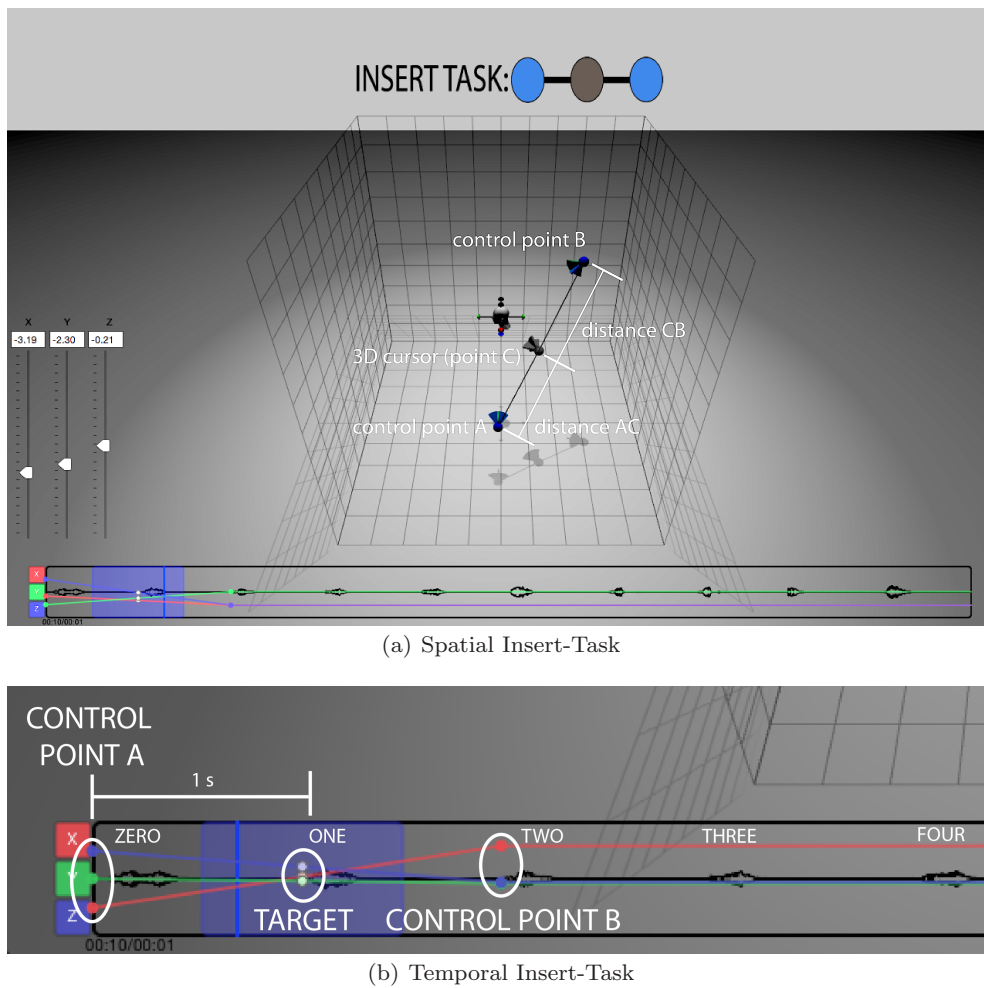


Fig. 3: Example of a spatial (a) and temporal (b) Insert-Task.

ers, GUI Timeline, Zonal Trackpad, Physical Sliders, and Phantom Controller.

#### GUI Sliders:

The primary input method was the trackpad as a pointing device to control the mouse cursor in order to interact with three spatially-multiplexed GUI sliders placed next to the 3D virtual scene (see Figure 5(a)).

#### GUI Timeline:

Designed to simulate standard automation line drawing in audio production software, this trackpad input method and cursor were used to control the three 2D control points along the GUI timeline. Each control point (red ‘x’, green ‘y’, and blue ‘z’) represented a Cartesian coordinate parameter of the 3D cursor and manipulated it in real-time within the 3D virtual scene. Participants can quickly change the operational mode between the three parameters via the keyboard shortcuts of ‘x’, ‘y’, and ‘z’ (see Figure 5(b)).

#### Zonal Trackpad:

This new interaction technique was inspired by Aceituno and Potier (2014) and used the multi-touch capability of the trackpad. This design spatially-multiplexed the surface of the trackpad by defining zones that are directly mapped to control the  $x$ ,  $y$ , and  $z$  parameters of an audio object. Figure 5(c) describes how touching the upper right zone designates that touch to control the  $x$  and  $y$  parameter, and the touch’s movement on the trackpad’s  $x$ -axis and  $y$ -axis was mapped to change the  $x$  and  $y$  parameter. The same touch method was used for the left zone, which was mapped to change the  $z$  parameter.

#### Physical Sliders:

This interaction technique used three physical sliders of an external MIDI controller (Akai APC Mini <sup>4</sup>) that were spatially multiplexed. This device used the MIDI

<sup>4</sup> <https://www.akaipro.com/apc-mini>



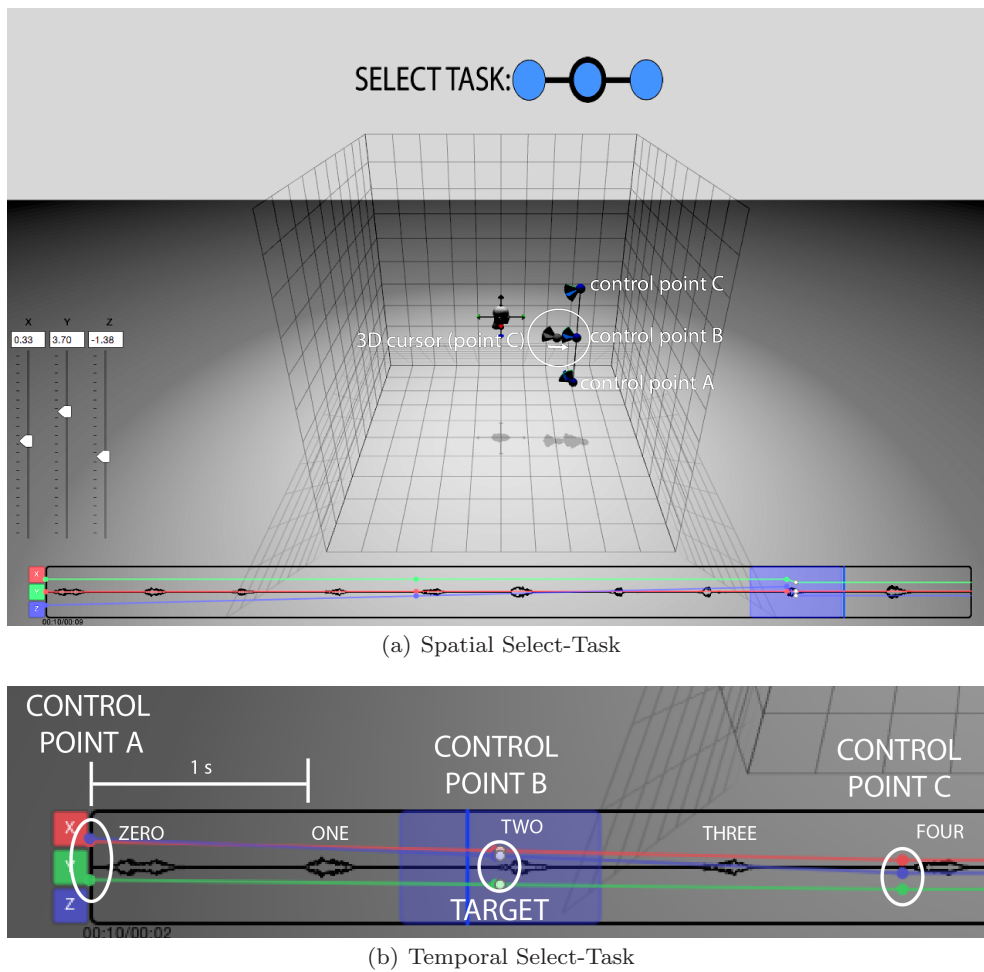


Fig. 4: Example of a (a) spatial and (b) temporal Select-Task.

protocol to send MIDI values (0-127) when the controller changes (rather than continuously sending values). The values of the slider are scaled to have a direct-absolute mapping to the range of the spatial parameters (see Figure 5(d)).

#### *Phantom Controller:*

The 3D positional input of a Phantom Desktop device was mapped to the range of the 3D positional parameters (see Figure 5(e)). The Phantom Desktop is a 6-DoF controller that provides a pen like grip and haptic force-feedback (now rebranded as 3D systems Touch X<sup>5</sup>). For this study, we only employed it as a 3D positional controller, adding only a small continuous resistance to help reduce the amount of jitter within the control mechanism of the device. No other haptic functionality of the device was employed.

The Zonal Trackpad, Physical Sliders, and Phantom Controller are interaction techniques that directly

manipulate the object of interest (the 3D cursor) by mapping the inputs to control the position of the 3D cursor. The GUI Sliders and GUI Timeline are indirect interaction techniques in that the user must directly control the operating system's cursor through a mouse and trackpad, and then position to cursor and use operations to control a set of GUI elements that are directly mapped to the positional parameters of the 3D cursor.

### 3.4 Temporal Interaction Techniques

Participants conducted the temporal tasks with three interaction techniques that used the standard laptop trackpad as the input device. These interaction techniques were mapped to control the temporal cursor and the position of the audio playback window. Two of the interaction techniques used the mouse cursor to interact with the timeline. A third interaction technique investigated the use of a discrete control (swipe gesture) to move the playback window. This discrete technique

<sup>5</sup> <https://www.3dsystems.com/haptics-devices/touch-x>

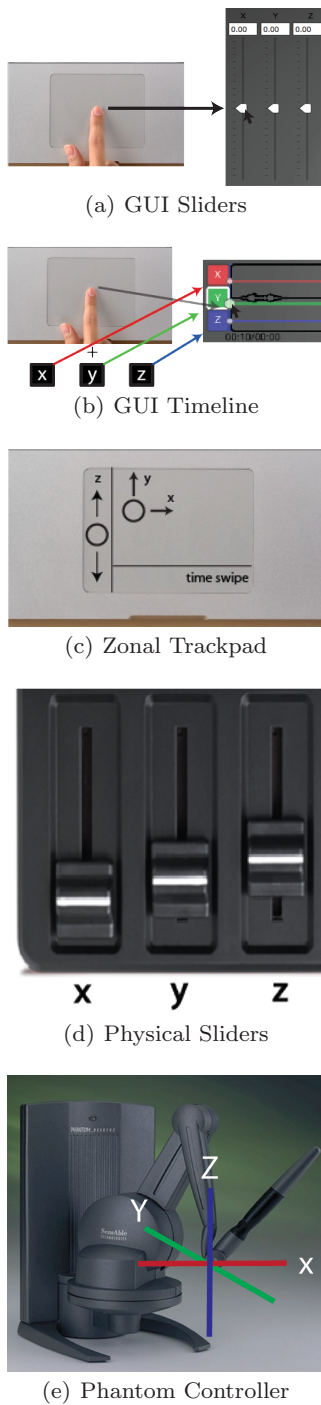


Fig. 5: Spatial interaction techniques.

was chosen in order to evaluate the use case of having small temporal controls (buttons, swipe, etc.) close to the spatial controls in order to reduce the spatial multiplexing of input methods. We labeled these techniques: *Cursor-Point & Click*, *Cursor-Continuous*, and *Swipe*.

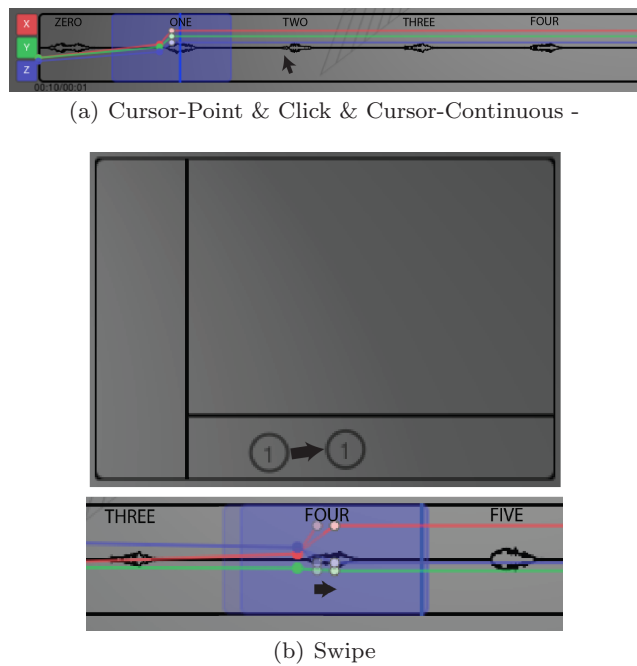


Fig. 6: Temporal interaction techniques that describe the *Cursor-Point & Click* and *Cursor-Continuous* using the mouse in the GUI timeline (a) and the discrete gestures used in the *Swipe* technique (b).

*Cursor-Point & Click:*

This standard interaction technique allowed users to point and click on the timeline using the mouse cursor. When users clicked on the timeline, the position of the temporal cursor and the audio playback window moved to the location of the click (see Figure 6(a)).

*Cursor-Continuous:*

The position of the temporal cursor and the playback window was mapped to the position of the mouse cursor within the timeline. Whenever the mouse was located or moved within the timeline, the temporal cursor and audio playback window moved instantly to the location of the mouse. This provided real-time change in the position of the temporal cursor and audio playback (see Figure 6(a)).

*Swipe:*

This is a distinctly different interaction technique that used the bottom zone of the Zonal Trackpad input method for a swipe gesture to navigate along the timeline (see Figure 6(b)). A swipe triggered movement of the center of the playback window by 0.1 s along the timeline. A left-swipe moved towards the beginning of the timeline while a right-swipe moved towards the end of the timeline. For the temporal *Select-Task*, we investigated this technique as a method to traverse through control

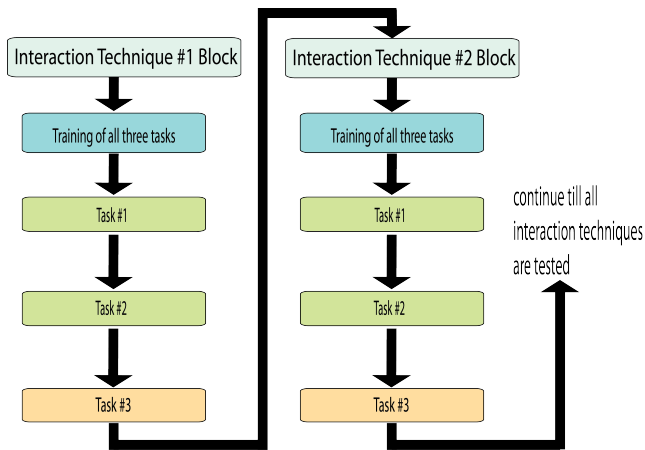


Fig. 7: Flowchart of experimental protocol during each session.

points. This was done by holding the Command-key along with a swipe, which moved the cursor to the closest existing control point in the direction of the swipe. In general, this is a different interaction technique in that the user has a smaller set of possible predefined positions in the timeline to position the cursor.

### 3.5 Protocol

Within each session, participants followed a protocol where they conducted 15 spatial conditions (5 spatial interaction techniques for 3 spatial tasks), 9 temporal conditions (3 temporal interaction techniques for 3 temporal tasks), with 10 repetitions per each condition. This protocol was divided into blocks for each interaction technique (five for spatial, three for temporal). Order of the interaction technique blocks was randomized for each participant to ensure counter-balancing. Within each interaction technique block, participants conducted a training phase to practice and learn the interaction technique for each task, which was followed by the experimental trials that were counter-balanced (see Figure 7).

Participants were instructed that the goal of each trial was finish the task as quickly as possible. Completion time of the tasks was measured from the start of the trial to when participants correctly validated their entered position within the spatial/temporal error tolerance of the target value, and qualitative data were recorded using surveys to measure performance, analyze behavior, and measure preferences. Surveys after each interaction technique block used Likert scale (1–5) questions to record impressions of the difficulty of the interaction technique in completing each task.

We also collected information on participants’ previous experience with 3D audio and 3D graphics (composing, developing, or basic experience) and their overall preferences within the spatial and temporal interaction techniques through a final survey. Observations of participant strategy and comments were recorded during the experiment. We report our quantitative analysis on completion time across tasks and interaction techniques by estimating the effect sizes with percent differences and confidence intervals using the bootstrap method (Cumming and Finch, 2005; Cumming, 2014; Kirby and Gerlanc, 2013).

## 4 Results

Twelve participants (9 men, 3 women) with experience in 3D audio or 3D graphics were recruited from research centers and universities. Eleven out of twelve participants had 5+ years of 3D graphic experience (92%) while two out of twelve had 5+ years of 3D audio experience (17%). Even though the participant pool is small, we observed similar ability to complete the tasks across these participants and between male and female participants. A small minority of participants had 3D audio experience, but their results did not significantly differ from the rest of the participants, presumably due to the instructions of using the visual feedback to monitor and validate the tasks. Completion time was analyzed by task and interaction technique by applying a log-transformation to compare effect sizes (Dragicevic, 2016). We present these analyses across spatial tasks, spatial interaction techniques, temporal tasks, and temporal interaction techniques through probability density distributions (histogram), averages (white text), and 95% confidence intervals (light gray bars). Subjective responses on the preferences of interaction techniques and the difficulty of use were gathered and analyzed using percentages and boxplot analysis respectively.

### 4.1 Results of Spatial Tasks & Interaction Techniques

Analysis across spatial tasks suggests an effect of the task’s instruction and structure on completion time. When comparing them together, participants completed the *Select-Task* (10.17 s) faster than the *Add-Task* (15.41 s) and the *Insert-Task* (12.51 s) with a 41% and 18% difference respectively. They conducted the *Insert-Task* faster than the *Add-Task* with a 24% difference (see Figure 8). These tasks had a similar general goal of moving the 3D cursor to a target location, but these results suggest that the differences in directions and

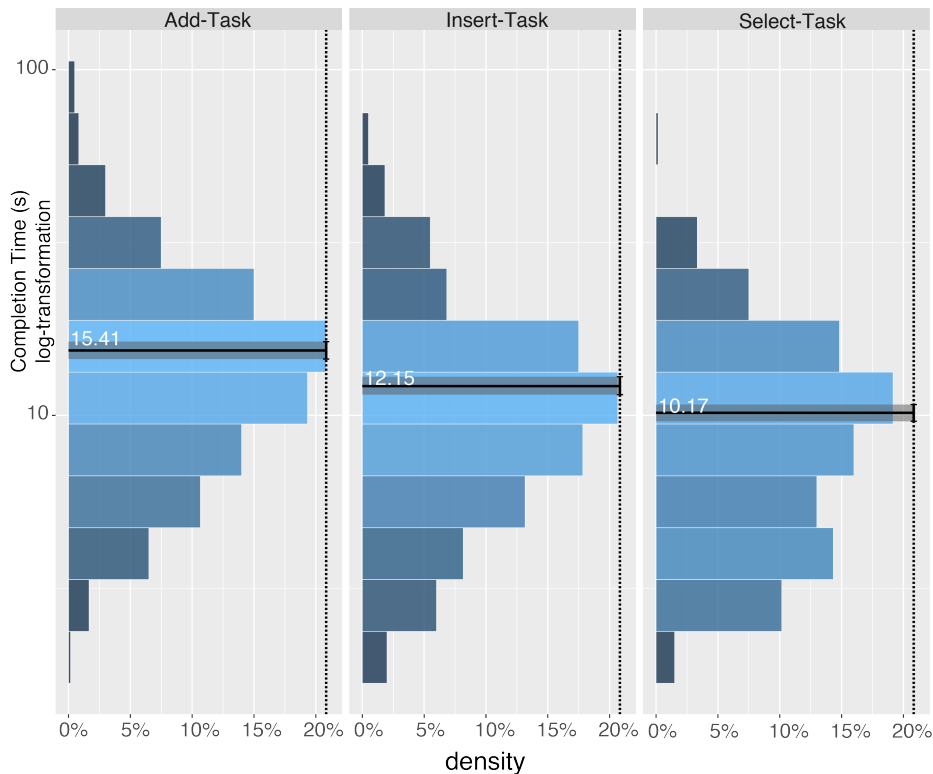


Fig. 8: Probability density distribution using log-transformation (histogram), mean averages (white text), and 95% confidence intervals (light gray bars) of completion time by spatial tasks.

structure of the task had an effect on completion time. This trend of participants completing the Select-Task fastest and the Add-Task slowest was also observed for each spatial interaction technique.

In Figure 9, the analysis of completion time across interaction techniques shows that participants completed the tasks fastest on average with the *Phantom* (6.56 s) and slowest with the *GUI Timeline* (21.31 s). Participants performed faster with the *Phantom* than the *textitGUI Sliders* (15.32 s) with a 80% difference and the *GUI Timeline* with a 106% difference. They performed with the *Zonal Trackpad* (8.14 s) faster than the *GUI Sliders* and *GUI Timeline* with a 61% and 90% difference respectively, and they completed the task faster with the *Physical Sliders* (11.5 s) than the *GUI Sliders* and *GUI Timeline* with a 28% and 59% difference respectively. These comparisons indicate that the direct manipulation design of the *Phantom*, *Physical Sliders*, and *Zonal Trackpad* designs performed better than the standard design of *GUI Slider* and *GUI Timeline*. When comparing the direct interaction techniques to each other, participants performed with the *Phantom* faster than the *Zonal Trackpad* and *Physical Sliders* with a 22% and 55% difference respectively. They performed with the *Zonal Trackpad* faster than the *Physi-*

*cal Sliders* with a 35% difference. This comparison supports the notion that a high degree of integrality in the interaction technique improves completion time in 3D positioning tasks, which is consistent with previous studies (Jacob et al, 1994; Martinet et al, 2012; Veit et al, 2009).

#### *Subjective Preferences:*

Participants were asked which interaction technique they preferred in completing each task. Table 1 reports results of preferred spatial interaction technique for each task. A strong preference was observed for direct interaction techniques (*Physical Sliders*, *Zonal Trackpad*, and *Phantom*), while no participant chose the *GUI Slider* or *GUI Timeline*.

Table 1: Preferred spatial interaction technique by spatial task across participants.

	Add	Insert	Select
GUI Sliders	0.0%	0.0%	0.0%
GUI Timeline	0.0%	0.0%	0.0%
Zonal Trackpad	50.0%	33.3%	25.0%
Physical Sliders	25.0%	50.0%	16.7%
Phantom	25.0%	16.7%	58.3%

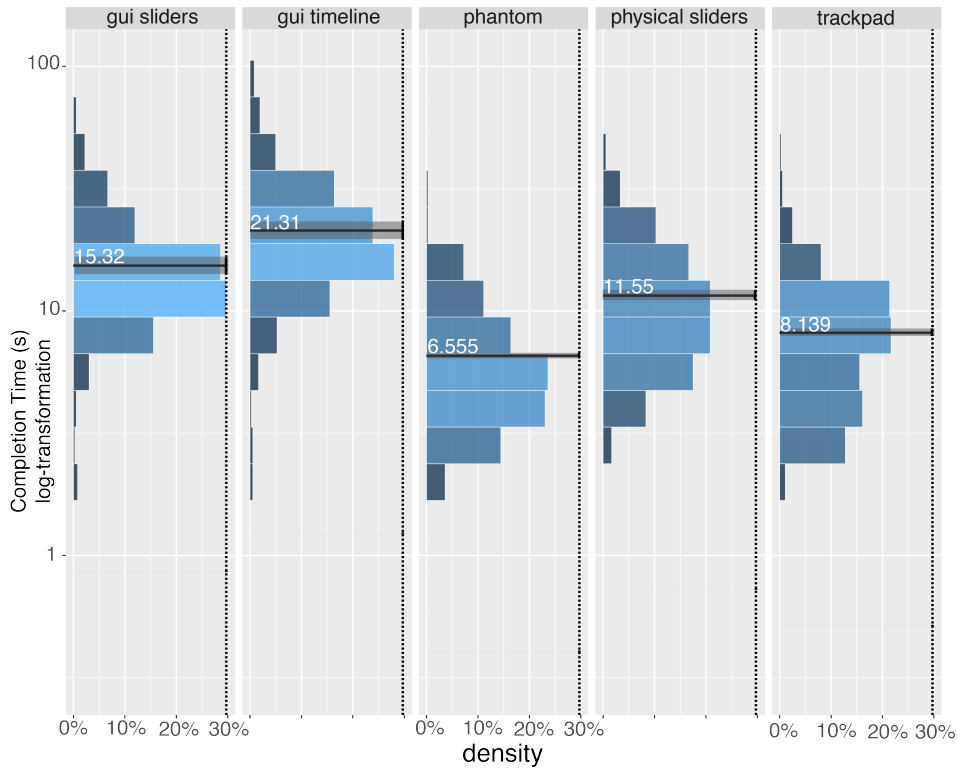


Fig. 9: Probability density distribution using log-transformation (histogram), mean averages (white text), and 95% confidence intervals (light gray bars) of completion time by interaction techniques.

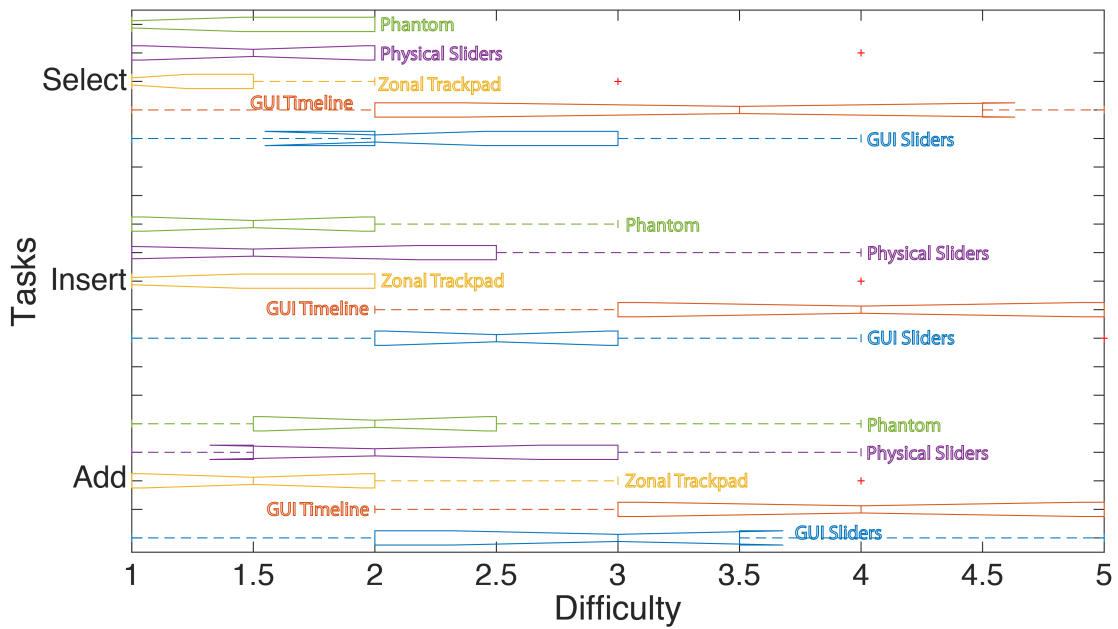


Fig. 10: Difficulty rating of GUI Sliders (blue), GUI Timeline (red), Zonal Trackpad (yellow), Physical Sliders (purple), and Phantom (green) interaction techniques within the Add-Task, Insert-Task, and Select-Task.

Similarly, participant responses on the difficulty to use the interaction techniques for each task showed clear

preferences for the *Phantom*, *Physical Slider*, and *Zonal*

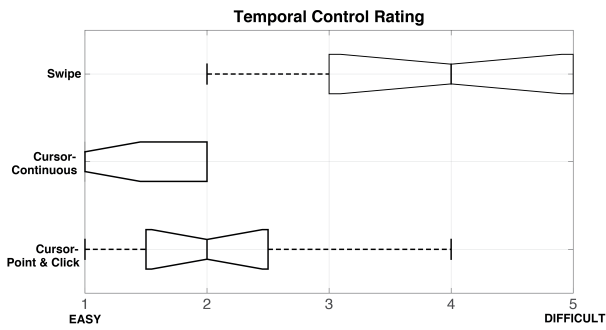


Fig. 11: Difficulty rating of each temporal interaction technique.

*Trackpad* as well. In these questions, participants rated the difficult in using each interaction technique according to a Likert Scale (1–5), with 1 indicating “Easy” and 5 indicating “Difficult”, for the *Add-Task*, *Insert-Task*, and *Select-Task* (see Figure 10). Results, along with completion time analyses, suggest that spatial interaction techniques that provide a direct manipulation were easier to learn and more preferred than the use of GUI elements, which is in line with previous work that investigated the level of directness of manipulation techniques (Aceituno, 2015; Shneiderman, 1997; Damaske, 1969).

#### 4.2 Results of Temporal Tasks & Interaction Techniques

Analysis of completion time across temporal tasks highlighted similar effects where participants completed the *Select-Task* (2.10 s) faster than the *Add-Task* (4.22 s) and *Insert-Task* (4.19 s) by a 67.1% and 66.5% difference respectively. However, there was little difference between the *Add-Task* and *Insert-Task* with a 0.7% difference.

When analyzing the tasks by each interaction technique, we observed little difference of average completion time between tasks when using the *Cursor-Point & Click* and *Cursor-Continuous* interaction techniques. Participants conducted the temporal *Select-Task* (1.98 s) faster on average compared to the *Add-Task* (2.64 s) and *Insert-Task* (2.01 s) by a 28.6% and 1.5% difference for the *Cursor-Point & Click* technique, and they performed slightly slower with the *Select-Task* (2.17 s) compared to the *Add-Task* (2.66 s) and *Insert-Task* (2.49 s) by a 20.3% and 13.7% difference for the *Cursor-Continuous* technique respectively. The differences between tasks for the *Swipe* interaction technique showed the *Select-Task* (2.14 s) was faster than the *Add-Task* (7.37 s) by 110.0% and *Insert-Task* (8.00 s) by a 115.6%

difference. The large differences between the average completion times in the *Swipe* technique were due to the design of the task and its discrete design. Participants used swipes to traverse through control points for the *Select-Task*, while they used swipes to navigate along the timeline for the *Add* and *Insert-Tasks*.

When comparing interaction techniques by task, participants performed similarly with the *Cursor-Point & Click* (2.64 s) and *Cursor-Continuous* (2.66 s) technique by a 0.75% difference for the *Add-Task*. They performed slower using the *Swipe* (7.37 s) technique than the *Cursor-Point & Click* and *Cursor-Continuous* technique by a 94.5% and 93.9% difference. Differences between average completion times for the *Insert-Task* were slightly larger when between the *Cursor-Point & Click* (2.01 s) with the *Cursor-Continuous* (2.49 s) with a 21% difference. They performed the same task slower with the *Swipe* (8.00 s) technique by a 119.6% (*Cursor-Point & Click*) and 105% (*Cursor-Continuous*) difference. However, we observed similar average completion times across interaction techniques for the *Select-Task*. Participants performed slightly faster with the *Cursor-Point & Click* (1.98 s) technique compared to the *Cursor-Continuous* (2.17 s) and *Swipe* (2.14 s) techniques by a 9.16% and 7.77% difference, which suggests that participants could use the *Swipe* technique with similar performance to the standard cursor techniques as a method to traverse through control points rather than using it to navigate incrementally along the timeline.

Participants performed slightly faster with the *Cursor-Point & Click* technique than with the *Cursor-Continuous* technique, but the majority indicated that the *Cursor-Continuous* technique (66.7%) was easiest to learn compared to the *Cursor-Point & Click* (33.3%) and *Swipe* (0%) techniques. In addition, they preferred the *Cursor-Continuous* technique (66.7%) compared to the *Cursor-Point & Click* (25%) and the *Swipe* (8.3%) techniques. They rated the ability to use the *Cursor-Continuous* technique to control the time parameter easier compared to the *Cursor-Point & Click* and *Swipe* technique (see Figure 11). This suggests including real-time feedback is preferred when locating temporal positions within an audio recording, even though it can be unintelligible when scrolling quickly across the GUI timeline and it does not improve completion time compared to the *Cursor-Point & Click* technique.

When comparing the differences between spatial and temporal tasks, results showed little differences between the temporal *Add-Task* and *Insert-Task* for each interaction technique. This suggests that the structure of the temporal *Add-Task* and *Insert-Task* are quite similar even though the tasks’ directions and structures were

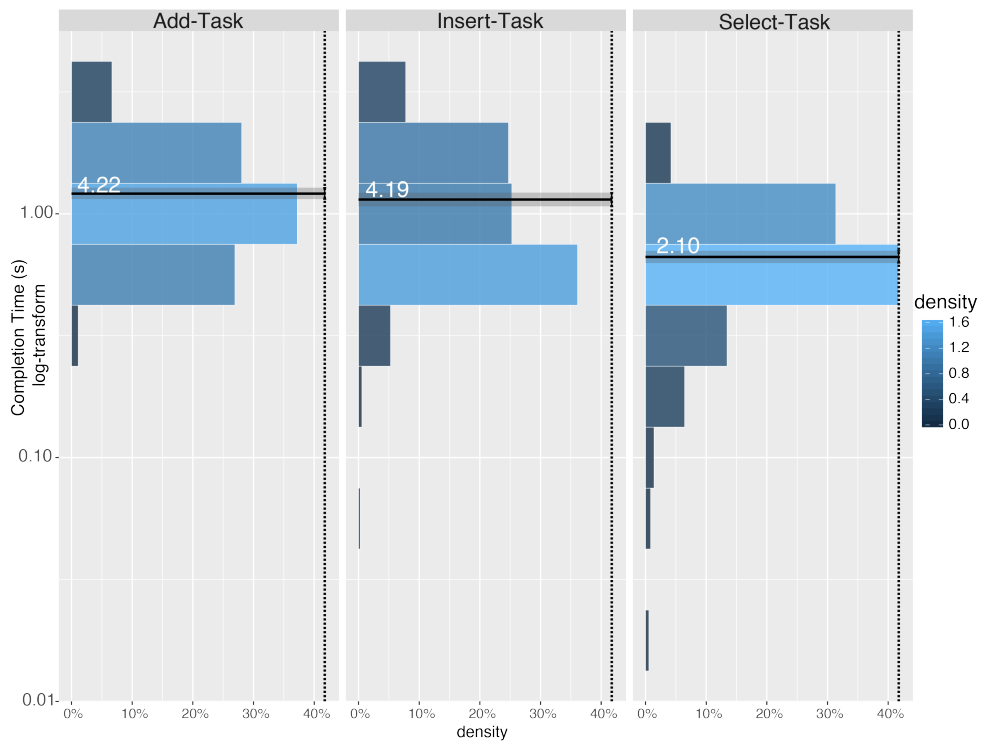


Fig. 12: Probability density distribution using log-transformation (histogram), mean averages (white text), and 95% confidence intervals (light gray bars) of completion time by temporal tasks.

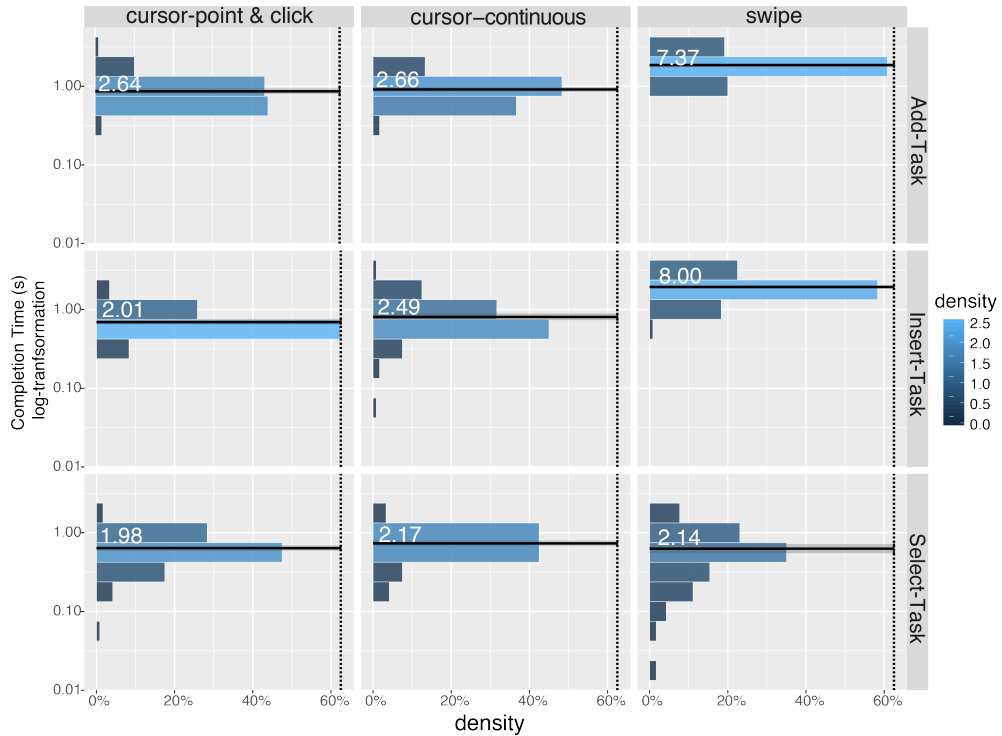


Fig. 13: Probability density distribution using log-transformation (histogram), mean averages (white text), and 95% confidence intervals (light gray bars) of completion time across temporal tasks by interaction technique.

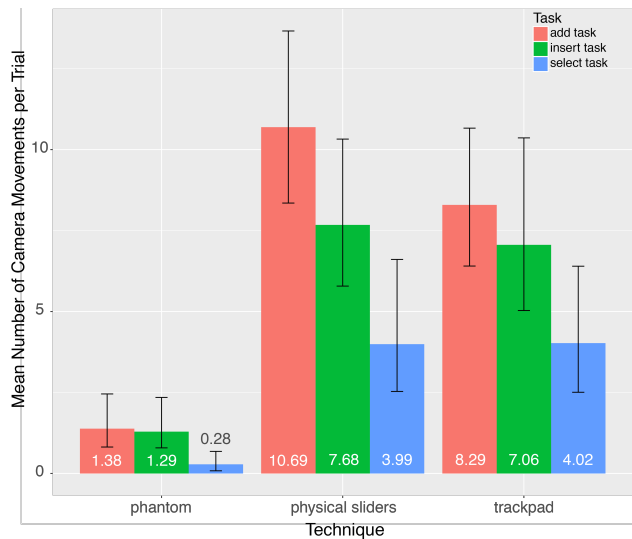


Fig. 14: Average number of camera movements per trial by technique as a function of task.

different. Both temporal tasks had participants search for a specific location within the timeline, and the target location did not greatly affect completion times. This was different to the spatial tasks’ effect on completion time where we observed larger differences between all three tasks. It is important to note that the temporal and spatial *Select-Task* serve equivalent purposes in a composite editing task, and the results between them indicate selecting control points was much faster using temporal interaction techniques on the timeline compared to spatial interaction techniques in the 3D scene by a 131.5% difference. This was expected due to the single dimensionality of interacting with the timeline compared to the three dimensions within the 3D graphics scene as there are less parameters to control.

#### 4.3 Observations on Integrality

In addition to the effect of integrality on completion time, we observed the average number of camera movements across participants were smaller with higher integrality of the device (see Figure 14). Across all three tasks, mean number of camera movements per trial reduced significantly with the *Phantom* (0.91) interaction technique compared to the *Physical Slider* (7.17) and *Zonal Trackpad* (6.20). This suggests that the structure of the *Select-Task* and the input method of the *Phantom* reduced the need for numerous camera movements, allowing participants to complete the tasks faster.

*Number of Combined Degrees of Freedom (NCD):*

Specific comments and observations indicated a differ-

ence between the integrality of interaction techniques (*Physical Sliders*, *Zonal Trackpad*, and *Phantom*) and that many participants completed the tasks using a strategy of separated operations rather than a single 3D operation. Some participants explicitly stated using this strategy. One participant stated that they “aligned then pushed towards the target” for the *Select-Task*, while another participant mentioned “2D + 1D controls is what I have been doing since the beginning” when commenting on the the *Zonal Trackpad*. Another participant mentioned that they only used “2 axes at once” for the *Physical Slider*, *Zonal Trackpad*, and *Phantom*. Some participants positioned the 3D cursor along a 2D plane first, and then adjusted the third dimension to complete the spatial tasks. This indicated that participants deconstructed the structure of the task as separate operations rather than an integrated operation as expected by the theory on integrality and separability Jacob et al (1994). However, the results of (Veit et al, 2011) indicated that the separability of interaction techniques is likely to improve precision of the 3D manipulation task.

To explore this in more detail, we conducted a Number of Combined Degrees of Freedom (NCD) analysis (Veit et al, 2011), which is an average percentage of time within each trial that participants interacted with any of seven possible combinations of Cartesian coordinate parameters (X, Y, Z, XY, XZ, YZ, XYZ). Figure 15 shows a NCD analysis of each input method with varying thresholds. Each threshold value determines at what point a dimension was considered engaged, when the change in that dimension exceeded the stated threshold.

For a zero (0) threshold, participants interacted with all three parameters (XYZ) the majority of time (94.94%) with the *Phantom* controller, while they interacted with the XY parameters 41.83% of the time with the *Zonal Trackpad*, and mostly with the X (32.28%), Y (36.49%), and Z (26.11%) parameters separately with the *Physical Sliders*.

For non-zero threshold values, changes that exceeded the associated threshold value within the dimension were considered as a user’s *intention* to interact with that dimension, above the noise of the interface. This investigated whether eliminating jitter in the analysis affected the assumption of user interaction with the input methods. When increasing threshold values, the NCD-XYZ value decreased from 94.9% (0.0 m) to 68.6% (0.0025 m), 57.7% (0.005 m), and 45.9% (0.01 m) for the *Phantom* input device. However, the NCD-XY value showed almost no difference with higher thresholds from 41.8% (0.0 m) to 41.5% (0.0025 m), 41.1% (0.005 m), and 40.4% (0.01 m) for the *Zonal Trackpad*. Also, the



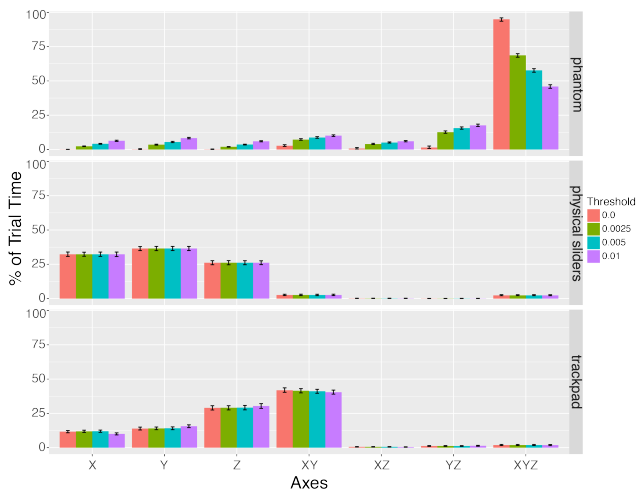


Fig. 15: NCD analysis of average time spent interacting with a combination of dimensions with threshold of 0 m (red), 0.0025 m (green), 0.005 m (blue), and 0.01 m (purple).

NCD-X, NCD-Y, NCD-Z values stayed consistent with the *Physical Sliders* with increased threshold values. This analysis highlights how separable input methods did not have large differences across thresholds and participants used them as separable interaction techniques even though they were able to bi-manually interact with them.

Combining this analysis with participant comments suggests they may have used the *Phantom* as a separable input method similar to that of the *Zonal Trackpad* rather than an integral device. This maybe due to how the conditions (locations of the given control points) were randomized, which may have caused the target position to be occluded at times. Due to the instructions of fast completion time and the requirement of a certain accuracy, participants most likely found the strategy of lining up the cursor using a view point of a certain plane first and then changing the view point to ensure the third dimension was in line to be faster and more precise than learning how to control all three dimensions at once when using the *Phantom*. This would likely be a different case if a more egocentric visual display was used, *e.g.* head-mounted displays with stereoscopic graphics rendering. Such an observation is more in line with the findings of Veit et al (2011), since participants were required to be precise to complete the task within the error tolerance condition. Also, participant experience with these interaction techniques and tasks would have an effect on whether they use the *Phantom* in a separable or integral manner. Users would be expected to use the device in an integral manner with more practice in similar 3D tasks, due to them learn-

ing the mapping of the controller’s position to 3D virtual space. However, more demographic information on experience with these software tools, controllers, and types of tasks are required to better understand the effect of these designs.

## 5 Discussion

As hypothesized, increasing the integrality of input devices and using direct manipulation reduced the average completion time for the spatial tasks. The perceptual structure of these input devices is distinctively different. The structure of the *Physical Sliders* was three sliders that are spatially-multiplexed along the same dimension. This differed with the *Zonal Trackpad* that provided two spatially-multiplexed zones that was designated to control along the  $X - Y$  plane and another designated for the  $Z$ -axis. Lastly, the *Phantom* controller was a single handed controller with 3 orthogonal axes that represented the Cartesian coordinate axes in the virtual scene. These results suggest that the perceptual structure of the spatial tasks were closely matched to the perceptual structure of the *Phantom* controller based on the theory of integrality and separability (Jacob et al, 1994).

It was also observed that the design of the visual references within the tasks affected the strategy of some participants to complete the tasks. The *Add-Task* provided a line from the 3D cursor to control point  $B$  (see Figure 2) to provide a visual reference to complete the task, while the *Insert-Task*’s structure provided two lines on either side of the 3D cursor to control point  $A$  and  $B$  (see Figure 3) for a visual reference. The *Select-Task*’s structure provided a target control point as a visual reference where participants were required to dock the 3D cursor (see Figure 4). These were used to help find the target location and use it as a reference to tell whether they positioned the cursor in the target location. For the spatial tasks, one participant noted the differences in visual references by commenting that “*Select-Task was easier to complete, maybe because there was a clear target*”, which was not the case for the *Add-Task* and *Insert-Task*. In addition, participants commented on the use of shadows of the 3D cursor and control points to help localize their positions in the 3D scene. One participant mentioned, “*the shadows helped a lot, but a z-axis shadow could help better perform the task*”. Even though shadows may have helped complete the majority of the tasks, instances of the cursor and control points being occluded required participants to change the camera for a different perspective of the 3D scene. One participant commented that it would be useful to have “*camera control within*

*the same [input] device*”, suggesting that camera control input methods need to be quickly accessed to complete these tasks faster and more accurately. The need to rely on these visual references and camera control are likely due to the use of 2D visual display. We suspect that stereoscopic displays that provide 3D visual immersion would likely differ participant responses and preferences.

Similar use of visual references in the temporal tasks was observed, which highlighted a visual bias in the design of these tasks. One participant commented, “*after learning the relationship between the visual waveform and the context of the audio, there was no need to listen to the audio*”. This might be a reason on why the use of real-time feedback did not have a large effect on completion time compared to not using real-time feedback. The reduction of dimensions to control in these tasks and the learning effect of the visual waveform likely made the average completion time across temporal tasks similar for those two interaction techniques. However, participants did have the ability to use the lines and control points in the GUI Timeline that represented the 3D trajectory as well. If they were not using the visual waveform, they could have been using the lines and control points or a combination of the two to complete the tasks. This visual bias was likely due to the goal of the temporal task in asking participants to find a target time and the GUI timeline having a temporal representation. Target times being used as the primary goal could be a realistic scenario, but another realistic scenario would be a 3D audio practitioner defining temporal changes in 3D trajectory based on the context of the audio. A redesign of the temporal tasks by randomizing the context of the audio for each trial and slightly changing the temporal Add, Insert, and Select tasks to force participants to listen to the audio would better evaluate this other scenario.

From the point of view of the composite 3D audio trajectory authoring task, this study was separated into temporal and spatial tasks and interaction techniques in order to compare the interaction techniques with less experimental factors. If these tasks and techniques were combined for a more general task, we suspect the average completion time would be larger than the sum of atomic tasks due to control mode switching with the same input device or switching between spatially-multiplexed controls. The added time required to complete the composite versus the sum of the atomic tasks has been described as completion time offset (Beaudouin-Lafon, 2000). Larger completion time offsets can disrupt the workflow of the user, which can negatively affect overall usability and make the interaction technique less preferred (Selker, 2005). Techniques such as keyboard

shortcuts are a common examples of solutions to reduce the time for switching control modes. When the cursor interaction techniques are paired with external input devices, users would have to switch between the spatially-multiplexed camera, trackpad, and the external input devices in this configuration. An augmented design could be applied to the Physical Sliders, Phantom, and other external input devices through the use of easily accessible buttons. For example, this could help reduce the completion time offset when acquiring temporal controls for the Select-Task. One issue that should be considered is selecting a target control point when a large number of control points are present, which would require a large number of actions when traversing through the different points. The cursor techniques or selection using 3D interaction may be more suitable in this scenario.

In general, this overall reflection and analysis provides a more informed methodology on how to combine atomic tasks into more complex composite tasks. We hypothesize that combining these atomic tasks would create composite tasks and interaction techniques that would require training in order to master. This would be seen in the results in the completion times of the tasks over time, where the average completion times of composite tasks would approach the sum of the average completion times of atomic tasks as participants master the task and interaction technique. However, further investigation and experiments in the atomic temporal tasks are required to understand the average completion time with a goal of a target context.

## 6 Conclusion

This work presented a classification of spatial and temporal atomic tasks (Add, Insert, and Select) that relate to the composite task of authoring a 3D audio trajectory. This classification describes the basic tasks to define and edit the path of a sound source’s movement and when those movements occur. With this, we conducted a user study that evaluated and analyzed participant ability to complete these tasks using different interaction techniques. Results of this study highlighted how the design of each spatial and temporal interaction technique across tasks affected completion time and user preference. In addition, participant feedback highlighted the importance of learning the use of available visual references, providing an explanation of results across the tasks and the visual bias with the design of the temporal tasks.

The results of this study extracted two major findings:

1. Increasing the integrality and direct manipulation of the spatial interaction techniques reduced completion times.
2. A discrete control, the swipe gesture, was found to be a suitable option to traverse a small number of control points.

In addition, we observed participants conducted the spatial Select-Task with lowest average completion time compared to the Add and Insert-Task. Further analysis of the results suggested the integral design of the spatial interaction techniques could have an effect on the user's need for camera and viewpoint manipulations. Observations suggested that participants occasionally separated the interaction when using a fully integral input device, while most of them used the separated input device as separated interactions even though they could use them bi-manually. In the context of temporal tasks, we observed that mapping the temporal cursor and audio playback window to the cursor had similar completion times to the use of a standard point and click interaction technique, even though participants indicated a preference towards having real-time change in the audio feedback. In addition, we observed the use of a target time in the design of the temporal task resulted in a visual bias with the timeline, allowing for task completion without use of the audio playback.

### 6.1 Future Directions

Other design characteristics, such as the tactile and haptic feedback, still need to be better understood for 3D audio trajectory authoring. This would require choosing different interaction techniques and evaluating this across the three tasks in order to properly evaluate the effect. Overall, this type of quantitative study would provide further understanding on the design's effect on workflow production, while allowing participants to experiment and explore each interaction design would provide useful information on their contribution to creativity.

New designs on the temporal tasks and interaction techniques to remove all visual bias would better evaluate the goal of finding a target context for audio. Tasks should vary the context of the audio in order to require participants to listen and search for a contextual part of the audio. Afterwards, future studies on how to combine temporal and spatial interaction techniques for composite tasks could provide results that could be applied to more practical 3D audio production systems. This would consist in more realistic tasks requiring participants to change multiple spatial control points, changing temporal positions of the control points, and in-

creasing the number of control points from an existing trajectory with random spatial and temporal properties and varying audio recordings. Combining discrete and continuous temporal interaction techniques with spatial interaction techniques would provide a more realistic evaluation scenario. This would oblige participants to switch between input methods or control modes in order to spatially and temporally compose and edit a trajectory, which would facilitate a proper evaluation of the effect of interaction technique design on overall usability.

The use of a 2D display is still the norm for most 3D audio production systems. Despite this, the use of advanced display technology, such as virtual reality displays or force-feedback, is another aspect that merits further investigation. Such change in visual feedback could increase the direct manipulation and integrality of interaction techniques by synchronizing 3D renderings from the visual and audio displays providing a constant egocentric point of view.

### 6.2 Summary

In summary, this work provided further insight and direction on methodologies for studying the design of user interfaces for 3D audio trajectory authoring. Results from both spatial and temporal tasks highlighted the expected effect of direct manipulation and integrality, and in turn identified a design methodology of 3D audio user interfaces that could improve workflow in 3D audio production tasks. As this study was designed using visual feedback as the main source of spatial information, these results can also be applied to the evaluation of interaction techniques for general 3D trajectory authoring for other types of media. However, a realistic 3D audio production scenario would require users to also evaluate the spatial information aurally and adjust positional parameters accordingly, requiring additional steps in the experimental design in order to understand the resolution of aural spatial listening experiences from a 3D audio display and rendering space. With a better description of the auditory localization resolution of the audio display, we could then define the associated error tolerance of the spatial tasks. Even though this would make the design of the study longer, evaluating these interaction techniques with only 3D audio feedback should help further understand their usability in a practical environment for 3D audio authoring tasks.

## References

- Aceituno J (2015) Direct and expressive interaction on the desktop: increasing granularity, extent, and dimensionality. PhD thesis, Université Lille 1 - Sciences et Technologies
- Aceituno J, Potier L (2014) The secret life of computers. In: CHI'14 Extended Abstracts on Human Factors in Computing Systems, ACM, pp 151–152
- Beaudouin-Lafon M (2000) Instrumental interaction: An interaction model for designing post-wimp user interfaces. In: SIGCHI Conf on Human Factors in Computing Systems, ACM, New York, NY, USA, CHI '00, pp 446–453
- Bergstrom T, Karahalios K (2007) Conversation clock: Visualizing audio patterns in co-located groups. In: System Sciences, 2007. HICSS 2007. 40th Annual Hawaii Intl Conf on, IEEE, pp 78–78
- Bresson J, Schumacher M (2011) Representation and interchange of sound spatialization data for compositional applications. In: Intl Computer Music Conf, pp 1–5
- Carpentier T (2015) Récents développements du Spatialisateur. In: Journées d'Informatique Musicale, Montréal, Canada, URL <https://hal.archives-ouvertes.fr/hal-01247502>
- Cumming G (2014) The new statistics: Why and how. *Psychological Science* 25(1):7–29
- Cumming G, Finch S (2005) Inference by eye: confidence intervals and how to read pictures of data. *American Psychologist* 60(2):170
- Damaske P (1969) Directional dependence of spectrum and correlation functions of the signals received at the ears. *Acta Acustica united with Acustica* 22(4):191–204
- De Araújo BR, Casiez G, Jorge JA (2012) Mockup builder: direct 3D modeling on and above the surface in a continuous interaction space. In: Graphics Interface, Canadian Information Processing Soc, pp 173–180
- Dragicevic P (2016) Fair statistical communication in HCI. In: Modern Statistical Methods for HCI, Springer, pp 291–330
- Dragicevic P, Huot S (2002) Spiraclock: a continuous and non-intrusive display for upcoming events. In: CHI'02 extended abstracts on Human factors in computing systems, ACM, pp 604–605
- Favory X, Garcia J, Bresson J (2015) Trajectoires: a mobile application for controlling and composing sound spatialization. In: 27ème conférence francophone sur l'Interaction Homme-Machine., ACM, Toulouse, France, pp 1–4
- Garcia J, Bresson J, Carpentier T (2015) Towards interactive authoring tools for composing spatialization. In: IEEE 10th Sym on 3D User Interfaces, Arles, France, pp 1–2
- Jacob RJK, Sibert LE, McFarlane DC, Mullen MP Jr (1994) Integrality and separability of input devices. *ACM Trans on Computer-Human Interaction* 1(1):3–26
- Jacquemin G, Coduys T, Ranc M (2012) Iannix 0.8. Actes des Journées d'Informatique Musicale (JIM 2012) pp 107–115
- Kirby KN, Gerlanc D (2013) Bootes: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods* 45(4):905–927
- Martinet A, Casiez G, Grisoni L (2010) The design and evaluation of 3D positioning techniques for multi-touch displays. In: 3D User Interfaces (3DUI), 2010 IEEE Sym on, pp 115–118
- Martinet A, Casiez G, Grisoni L (2012) Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. *IEEE Trans on Visualization and Computer Graphics* 18(3):369–380
- Mathew JD (2017) A design framework for user interfaces of 3D audio production tools. Thesis, Université Paris-Saclay, URL <https://tel.archives-ouvertes.fr/tel-01632727>
- Mathew JD, Huot S, Katz BFG (2017) Survey and implications for the design of new 3D audio production and authoring tools. *J on Multimodal User Interfaces* pp 1–11
- Melchior F, Pike C, Brooks M, Grace S (2013) On the use of a haptic feedback device for sound source control in spatial audio systems. In: Aud Eng Soc Conv 134
- Metatla O, Martin F, Parkinson A, Bryan-Kinns N, Stockman T, Tanaka A (2016) Audio-haptic interfaces for digital audio workstations. *J on Multimodal User Interfaces* pp 1–12
- Nielsen J (1993) Usability Engineering. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA
- Sachs E, Roberts A, Stoops D (1991) 3-draw: A tool for designing 3D shapes. *IEEE Computer Graphics & Applications* pp 18–25
- Schumacher M, Bresson J (2010) Spatial sound synthesis in computer-aided composition. *Organised Sound* 15:271–289
- Selker T (2005) Fostering motivation and creativity for computer users. *Intl J Human-Computer Studies* 63(4):410–421
- Shneiderman B (1997) Direct manipulation for comprehensible, predictable and controllable user interfaces. In: 2nd Intl Conf on Intelligent User Interfaces, ACM, pp 33–39

- Van Emmerik MJ (1990) A direct manipulation technique for specifying 3d object transformations with a 2d input device. In: *Computer Graphics Forum*, Wiley Online Library, vol 9, pp 355–361
- Veit M, Capobianco A, Bechmann D (2009) Influence of degrees of freedom’s manipulation on performances during orientation tasks in virtual reality environments. In: *16th ACM Sym on Virtual Reality Software and Technology*, pp 51–58
- Veit M, Capobianco A, Bechmann D (2011) An experimental analysis of the impact of touch screen interaction techniques for 3-d positioning tasks. In: *Virtual Reality Conf (VR)*, IEEE, pp 75–82
- Zelevnik RC, Herndon KP, Hughes JF (2007) Sketch: an interface for sketching 3D scenes. In: *ACM SIGGRAPH 2007 courses*, p 19
- Zhai S, Milgram P (1998) Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In: *SIGCHI Conf on Human factors in Comp Sys*, pp 320–327