AUDIO-VISUAL ANALYTICS OF GEOSCIENTIFIC DATA WITH TIMBRAL VARIATIONS AND IMMERSIVE INTERACTIONS

Shamima Yasmin

Computer Science and Electrical Engineering
Eastern Washington University
Spokane, WA 99202, United States
syasmin@ewu.edu

Chowdhury Qaizar Jamal

Computer Science
University of Waterloo
Ontario, ON N2L 3G1, Canada
cjamal@uwaterloo.edu

ABSTRACT

This research investigated audio-visual analytics of geoscientific data in virtual reality (VR)-enhanced implementation, where users interacted with the dataset with a VR controller and a haptic device. Each interface allowed users to explore rock minerals in unimodal and multimodal virtual environments (VE). In the unimodal version, color variations demonstrated differences in minerals. As users navigated the data using different interfaces, visualization options could be switched between the original geographical topology and its color-coded version, signifying underlying minerals. During the multimodal navigation of the dataset, in addition to the visual feedback, an auditory display was performed by playing a musical tone in different timbres. For example, ten underlying minerals in the sample were explored. Among them, anorthite was represented by nylon guitar, the grand piano was used for albite, and so on. Initial findings showed that users preferred the audio-visual exploration of geoscientific data over the visual-only version. Virtual touch enhanced the user experience while interacting with the data.

1. INTRODUCTION

In terms of presence and user preference, multisensory systems have been proven superior to traditional virtual systems based on visual feedback [1]. Virtual reality (VR) is widely used in simulations and visualization applications to make the overall user experience more immersive and engaging [2]. Research findings demonstrated that VR provided a more intuitive visualization experience within a synthetic space to help users understand the complexity of the data dynamics, including both geographical and urban environmental data [3, 4, 5]. In addition to a wider field of view to match the real-life experience, VR provided sophisticated sound spatialization with binaural audio, creating a 3D sound effect [2]. Most saliently, Hunt et al. [6] emphasized interaction with sonification to promote user engagement with the virtual environment (VE) and achieve a fluent interaction style. In addition, Correia et al. [7] found a congruent audio-visual display resulted in better performance and higher engagement than arbitrary associations between sound and image. Transforming data into sounds for auditory display provided new perspectives for analyzing and interpreting scientific data, such as ways to map protein molecules to music [8]. Multimodality allows users to integrate information at an enhanced rate because of the associative connections multiple stimuli provide the brain. Geological terrains are characterized by undulated topologies, with wide variations in rocks and minerals, all of which, with audio-visual representations and VR enhancement, would allow users to absorb more in-depth information from the model.

Most VR applications emphasize the expansiveness of the virtual world and the scope for interactivity, whereas this research focuses more on multimodality, which has yet to be explored to its fullest in VR-enhanced visualization. This research investigated multimodal analytics of geoscientific data in a VR-enhanced implementation, where VR controllers and a haptic device were used to interact with the dataset. Each interface allowed users to explore underlying minerals in unimodal (visual) and multimodal (audio-visual) VEs. In addition, the haptic device interface allowed users to touch and feel the data. In the unimodal version, in addition to the original geographical texture, a color-coded version demonstrated ten underlying minerals in different colors to understand their concentrations across the data visually. As users navigated the data using different interfaces, visualization options could be switched between the original geographical topology and its color-coded version. In the audio-visual mode, an auditory display was performed in addition to the visual feedback. A musical tone was played by ten different instruments for the auditory display of ten minerals. Initial findings showed that users preferred the audiovisual exploration of geoscientific data with the haptic device compared to other options. The timbral display provided distinct audio feedback for minerals. The experience of virtual touch with the haptic device interface enhanced user experience as they interacted with the data.

2. BACKGROUND

Multimodality can improve task-specific human performance in various contexts [9, 10]. Sighted users' visualization experiences were enhanced with sonification [11]. Scientific visualization, where data variations are highly irregular to differentiate visually, sonification improved user perception [12]. Research findings demonstrated that visual learning materials augmented with audio feedback enriched learners' experiences [13]. Sonification facilitated visual perception and helped users overcome challenges in visual representations [11]. The human brain is naturally wired to combine different modalities into a unique perception while interacting with the real world [14]. The ability to touch, feel, and grab provides better controllability over virtual objects, increasing the sense of presence and awareness. Hence, cross-modality influenced the overall user perception [15, 16, 17]. According to Rosli and Cabrera [18], the sheer amount of data for visual display could be challenging for visual perception. Combining visual and auditory displays could be more effective than unimodal visual demonstrations.

In VR, multimodality facilitates feelings of immersion and presence [19]. VR integration adds richness to the VE. The virtual world becomes more expressive through sight and sound. In an immersive VE, additional senses enhance the viewing experience [15, 16, 17]. VR has been sporadically used in interactive data visualization [20, 21]. Scientific visualizations, such as computational fluid dynamics simulation or geological topology exploration, improved user experience in VR[22, 23]. Heatmaps are widely used for geographical data visualization [24], however, audio-visual cartography with an immersive VE enriched user experience [25].

As Ben Shneiderman said, data is not just about presenting information; it is more about exploration through interactions [26]. Visual analytics allows users to analyze datasets using visual representations and identify patterns [27]. However, the " Data Jazz" could be boring if overloaded with information [28]. So, finding different ways to interact with the data has become imperative. As discussed, audio-visual analytics would engage users with the datasets better during the exploration journey. VR provides the scope for immersive analytics, allowing users to freely interact with the datasets, which is impossible in a desktop-based representation.

This research uses visual and audio-visual modes to explore geoscientific data in an immersive VE. Users interacted with the data in each mode with VR controllers and a haptic device. In the unimodal version, color variations demonstrated the presence of different minerals. A look-up table was used to map mineral types to timbres for the auditory display. Initial findings showed that users preferred multimodal exploration of terrain models. Virtual touch enriched user experiences. The timbral display scaled well with minerals to provide distinct audio feedback to users as they navigated the dataset using two different interfaces.

3. THE PROPOSED APPROACH

The proposed approach investigated variations in mineral rocks in terrain in visual and audio-visual modes in a VR-enhanced VE. It comprises the following components: visual representation, auditory display, and user interactions with different interfaces.

3.1. Visual Representation

Since the late 1980s and early 1990s, volcanic lava from Mt St Helens, Washington, United States, had been extensively studied, and various minerals were identified (https://www.rockcollector.co.uk/editorial1004.htm). A 3D model of Mt St Helens was chosen as the dataset. Figure 1a shows the top view of Mt. St. Helens in its original texture. For the preliminary study, the model was colored differently to visually demonstrate the presence of different minerals in rocks, as shown in Figure 1b. Ten different minerals were represented in ten different colors. These are albite, quartz, anorthite, orthoclase, magnetite, enstatite, ferrosilite, ilmenite, diopside, and hedenbergite, which were colored red, blue, magenta, cyan, yellow, green, grey, olive, black, and pink, respectively (Figure 1c). Concentrations of minerals determined the terrain color. For example, if the concentration of quartz was higher than other minerals in a particular area, that part was colored blue. Hence, there are two visualization options: the model with the original texture and its color-coded version representing underlying minerals.

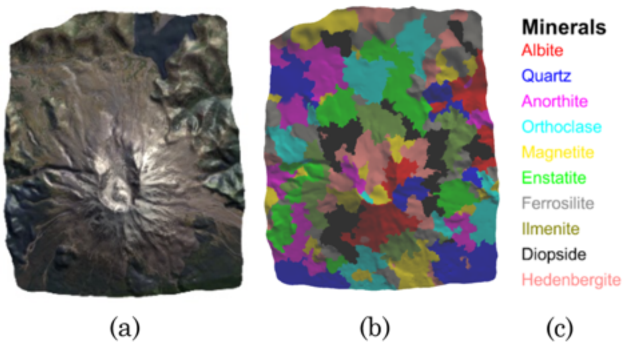

Figure 1: (a) Top view of the terrain in its original texture, (b) the terrain model is color-coded demonstrating underlying rock minerals, and (c) different minerals present in the terrain.

3.2. Auditory Display

Sonification is the use of non-speech audio that translates data into sound. Research findings demonstrated that sonification-based data representations could engage people emotionally and had the advantage of a deeper and richer understanding of data variations [29]. Users with higher musicality exhibited higher accuracy in interpreting sonified data tables [11]. In addition, popular music can help novice users understand subtle tonal differences produced by variations in auditory parameters. The auditory display was performed via sonification. Among auditory parameters, the pitch was reported as the most intuitive [30]. The pitch, a comparative high-low measure of sound changing logarithmically with frequency, can be varied within a narrow range to produce distinct output.

Stephen Barrass [31] demonstrated that timbre can be used as a complete (circular) dimension. Timbre allows us to distinguish the unique sound qualities of different musical instruments [32]. Ten musical instruments played a particular tone for the auditory display of ten minerals. These are grand piano, harpsichord, nylon guitar, sitar, viola, cello, baritone saxophone, bassoon, accordion, and shakuhachi. Thus, each mineral has its visual and audio-based representations. Table 1 shows colors and musical instruments used to demonstrate the visual and audio-based representations of different minerals in Mt. St. Helens.

Three different types of musical instruments represented different colors. These are stringed instruments, reed instruments, and flutes. Stringed instruments produce sound by stretching strings, while reed instruments have metal reeds, which produce sound when air is blown through them. String instruments, i.e., grand piano, harpsichord, nylon guitar, sitar, viola, and cello, were used for the timbre-based auditory display, produce bright and sharp sounds and were matched with bright colors, i.e., red, blue, and magenta, cyan, yellow, and green, respectively. The corresponding minerals were albite, quartz, anorthite, orthoclase, magnetite, and enstatite. Reed instruments, such as baritone saxophone, bassoon, and accordion, generate loud and heavy sounds

Table 1: Visual and auditory displays of minerals through colors and timbres.

Mineral	Color	Timbre
Albite		Grand Piano
Quartz		Harpsichord
Anorthite		Nylon Guitar
Orthoclase		Sitar
Magnetite		Viola
Enstatite		Cello
Ferrosilite		Baritone Saxophone
Ilmenite		Bassoon
Diopside		Accordion
Hedenbergite		Shakuhachi

and were matched with dull colors such as grey, olive, and black. The corresponding minerals were ferrosilite, ilmenite, and diopside. Flutes are the only woodwind instruments that do not use a reed; the musician has to blow in its tone holes across a lip plate. Flutes like Shakuhachi produce a soft tone; the corresponding color was pink to represent hedenbergite.

Figure 2 shows part of the model in the wireframe. For the audio-visual mapping, as discussed, each vertex had a timbral representation depending on its color. In the figure, grey represented ferrosilite, and the baritone saxophone was used for the timbral display. Similarly, magenta represented anorthite, and nylon guitar was used as the corresponding timbre.

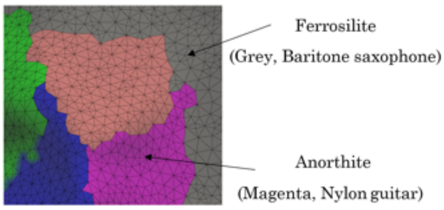

Figure 2: Each vertex of the terrain model is assigned color and timbre.

3.3. User Interface

Each mode (visual and audio-visual) was explored with two interfaces: VR controllers and a "Touch" haptic device. In an immersive VE, users put on VR headsets and interacted with the terrain model with VR controllers and a haptic device. Figure 3 shows the device setup, different buttons on a VR controller, and the Touch haptic device handle.

When pressed down, the gripper button in a VR controller would let users grab a model. Similarly, by default, the first stylus button of the haptic device was pressed to grab a model in a VE. As these buttons were released, objects were no longer being grabbed and left in the scene with the last transformations (i.e., translation and rotation) before being released in the VE. With VR controllers, interaction and exploration were performed by casting a ray onto the model by pressing the controller's trigger button (Figure 4a). For the haptic device, as users moved the handle, a cursor in the VE allowed them to touch and feel virtual objects (Figure 4b). These are detailed in the implementation section.

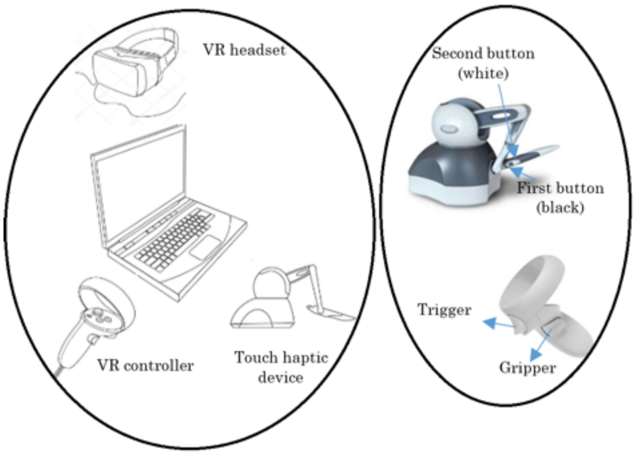

Figure 3: (Left) Different user interfaces (UI) for interactions in a VE and (right) different buttons on a haptic device handle and a VR controller.

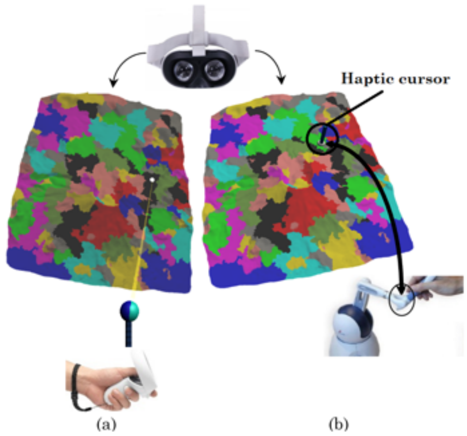

Figure 4: User interaction with a terrain model with different interfaces: (a) With a VR controller, users cast a ray to interact with the model. (b) A haptic cursor resembling the handle of the haptic device directly touches the model.

4. IMPLEMENTATION AND RESULTS

The project used Unity3d for its implementation. The Unity interface provides an easy plug-in for haptic devices so users can touch and feel virtual objects. Unity's XR plug-in option allowed users to explore models in VR. For auditory display, mp3 audio files were used for timbres from different musical instruments.

With a VR controller in the vision-only mode, the user could grab the model, move it around the scene, and bring it near or far to explore it. Users cast a ray onto the model by pressing the trigger button of the controller. The vertex closest to the ray-model intersection point is shown in white, and the corresponding mineral name (Albite) is highlighted on the screen (Figure 5 a). Users could switch between the original textured model and its color-coded version by toggling the "Show Materials" button. The "Show Materials" button turned yellow as the model's color-coded version was displayed. This is shown in Figure 5 b. The user

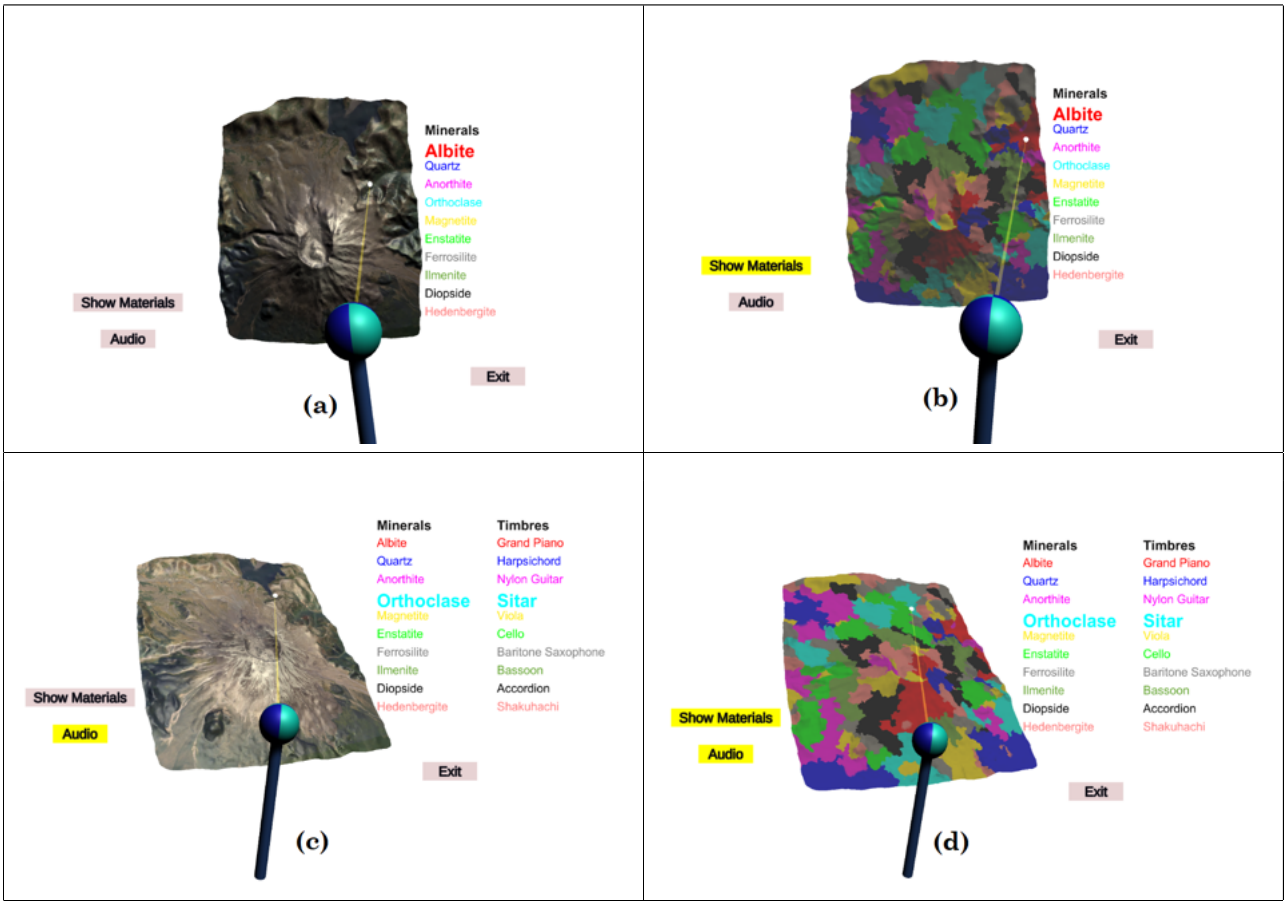

Figure 5: User interaction with the VR controller: (a) The user interacts with the terrain model with a VR controller (shown as a blue handle in the scene) by casting a ray in the default visual mode, i.e., the model in its original texture. The corresponding mineral (albite) on the screen is highlighted. (b) In visual mode, when the "Show Materials" button is turned on, the color-coded version of the model demonstrating colors for different minerals is shown. (c) The user interacts with the model in the audio-visual mode by turning the "Audio" button on. The model is shown in its original textured version. The corresponding timbre (sitar) and the mineral name (orthoclase) are highlighted on the screen. (d) The user interacts with the model in the audio-visual mode with the "Show Materials" button turned on.

toggled between the "visual" and "audio-visual" modes by toggling the "Audio" button. The "Audio" button toggled when the user cast a ray onto it with the simultaneous press of the trigger button. The "Audio" button turned yellow when the audio-visual mode was active. Names of musical instruments producing varying timbres were displayed beside the mineral names. The mineral name and corresponding timbre were highlighted as the user interacted with the model. In the figure, orthoclase was highlighted with sitar. In the audio-visual mode, as a ray was cast onto the model by pressing the trigger button, the model responded with the timbre assigned to the closest vertex. The vertex closest to the ray-model intersection point is shown in white in Figure 5c. As the user continued interacting with the model by casting rays, the model responded with different timbres determined by the ray model intersection points. In the audio-visual mode, users could switch between the original textured and color-coded models by toggling off and on the "Show Materials" button. This is demonstrated in Figure 5c and Figure 5d, respectively.

The virtual workspace was mapped to the device space to interact with the model with a haptic device. This allowed users to touch virtual objects by moving the device handle. The haptic cursor resembling the movements of the device handle in the virtual world is shown as a small blue handle in Figure 6a through Figure 6d. Like the VR controller, "Show Materials" and "Audio" buttons could be toggled on and off by touching and pressing them with the first button on the device handle shown in Figure 3. The corresponding minerals were highlighted as the haptic cursor touched the model. In the audio-visual mode, the model responded with audio feedback as the haptic cursor touched it. Like the VR controller, the audio feedback was determined by the timbre of the vertex closest to the haptic cursor. User interactions with the terrain with the haptic device interface for the visual and audio-visual modes are shown in Figure 6a through Figure 6d.

5. EVALUATION

Eleven sighted individuals aged 20 to 55 volunteered to participate in the user study. A "Touch" haptic device shown in Figure 3

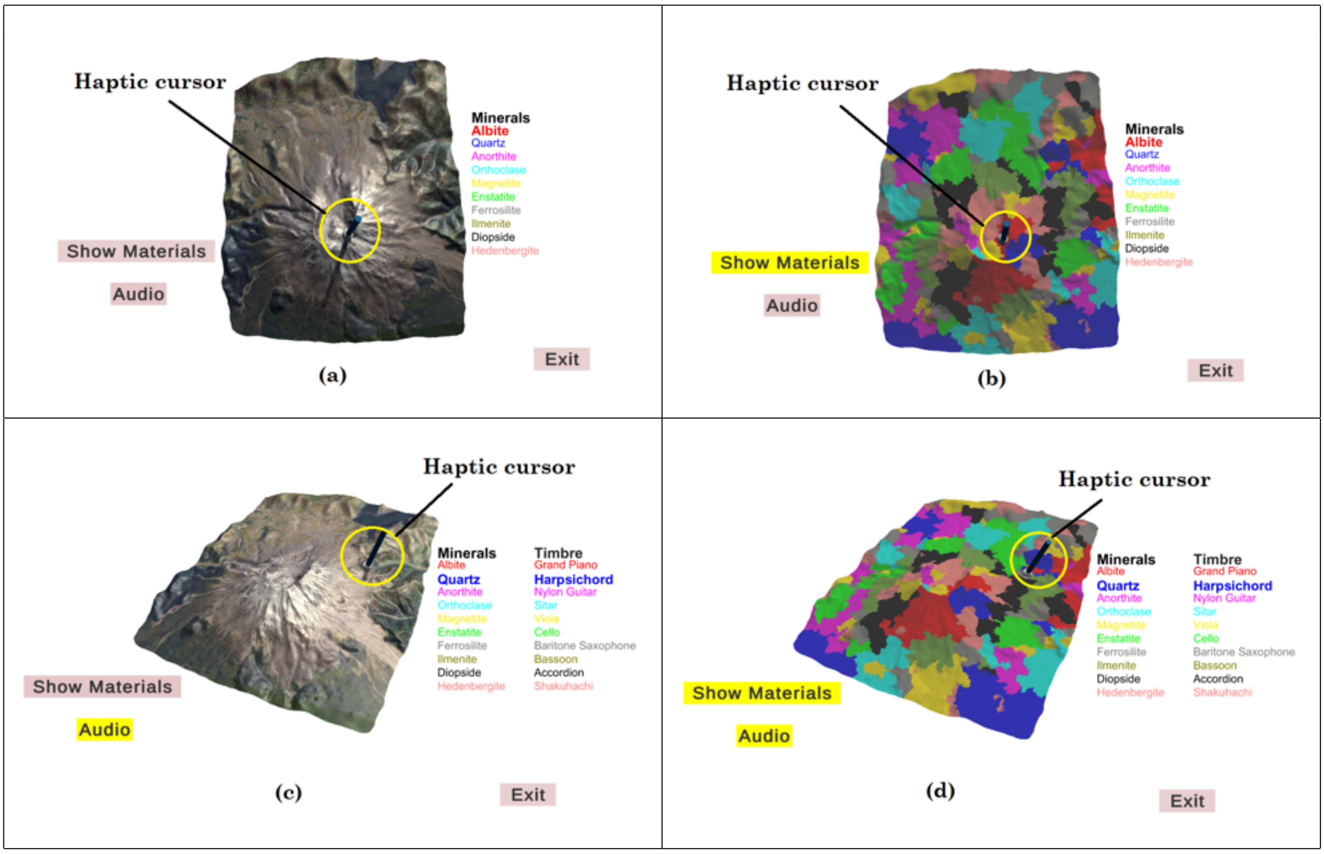

Figure 6: User interaction with the haptic device: (a) The user interacts with the terrain model by touching it with the haptic cursor (shown as a blue handle) in the default visual mode, i.e., the model in its original texture. The corresponding mineral on the screen is highlighted. (b) In the visual mode, as the "Show Materials" button is turned on, the color-coded version of the model demonstrating colors for different minerals is shown. (c) The user interacts with the model in the audio-visual mode by turning the "Audio" button on. The model is shown in its original textured version. The corresponding timbre, along with the mineral name, is highlighted on the screen. (d) The user interacts with the model in the audio-visual mode with the "Show Materials" button turned on.

and Figure 4 was used for haptic interactions. An Oculus Quest 2 headset was used for the VR integration. All participants were familiar with VR headsets. Three participants had prior experience with the "Touch" haptic device interface.

5.1. Study design and procedure

Recent research demonstrated participants stated performing tasks faster in the audiovisual mode compared to the vision-only option, though the measured response time was longer in the multimodal display compared to the vision-only version [11]. Wang et al. [29] ignored time but incorporated emotion and engagement to characterize the value of visualization. Considering the overall aspect of multisensory visualization, subjective measures were emphasized. The following metrics were considered while evaluating the usability with different modes and interfaces: user control, recognition, consistency, simplicity and aesthetic integrity, and accessibility (Figure 7) (`https://improvement.stanford.edu/resources/usability-principles`). Different evaluation metrics were briefly explained to users. For example, user control emphasized the interface's ease of interaction, allowing users to complete tasks efficiently and comfortably. Hence, users were asked to choose the option (i.e., the combination of mode and interface) that worked best for them. Recognition was about extracting the necessary information during navigation while switching between interfaces and modes. A consistent interface communicates clearly and efficiently with users; texts, labels, menus, and buttons look self-explanatory, clear, and concise. A simple and attractive interface allows users to focus on their work without causing distractions. Lastly, an accessible interface provides multiple ways to navigate, interact, and understand the data. Evaluation metrics were measured using a Likert scale ranging from 1 (poor) to 5 (excellent).

Subjects were asked to explore the terrain and underlying minerals using different interfaces and modes. They were first asked to gain an overview of the entire dataset before zooming in on different areas. As they zoomed in on a particular mineral, they were asked to identify neighboring minerals and clusters to find out relationships among minerals. For example, in our pilot study, quartz and anorthite, orthoclase and enstatite, diopside and hedenbergite created several clusters in pairs.

With the VR controller, participants explored the terrain in vi-

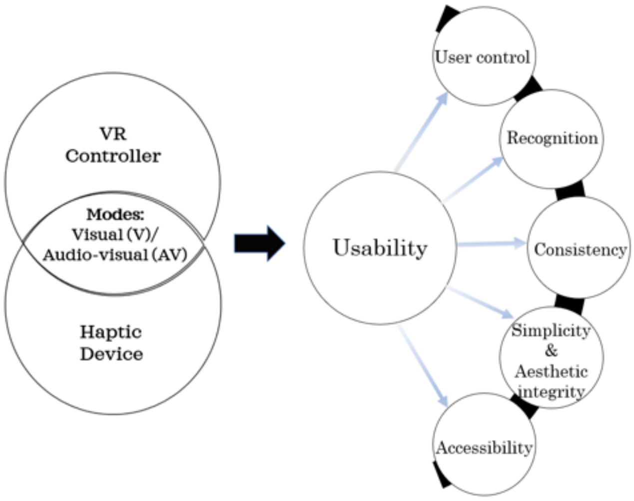

Figure 7: Different evaluation metrics to measure usability for different modes and interfaces.

sual and audio-visual modes by toggling the 'Audio' button. They did the same with the haptic device. The haptic device allowed them to touch and feel the model; no differences in surface properties were perceived. Hence, users evaluated different evaluation metrics for the following modes: (a) VR controller (visual), (b) VR controller (audio-visual), (c) haptic device (visual), and (d) haptic device (audio-visual). Here, the independent variable is the mode, and the dependent variable is the user evaluation score for different evaluation metrics. Participants put on the Oculus 2 VR headset for both interfaces. They were briefly introduced to devices and their tasks. Each interface was used separately without accessing the other. For example, with VR controllers, all interactions with objects or menus were performed with that interface. Similarly, a haptic device was used to perform the same tasks. Participants were given 30 minutes to explore the models in different modes. At the end of the study, they completed the Likert scale-based survey questionnaires. Participants were encouraged to provide comments and suggestions.

5.2. Results and analysis

Likert scale datasets are considered ordinal or ranked. As the normality of data was not confirmed for all options in the results, the non-parametric Friedman test was conducted to determine significant differences among the four options. Table 2 shows the Friedman test summary for usability. Here are the Friedman statistic (χ^2) and p-value: χ^2 = 25.48, p = 0.0000122 $<$ 0.0001. There were significant variations in subjective ratings with different modes.

Table 2: Friedman test summary for usability of the application and interfaces

Friedman statistic	P value	Kendall's coeff.	Effect size
25.48	0.0000122	0.77	Large

The post hoc comparison was conducted using the Durbin-Conover pairwise comparison test with Holm corrections. The user evaluation scores varied significantly with modes. The bar charts in Figure 8 demonstrate variations in usability scores with modes. Asterisks in the bar graphs showed significant differences between modes, i.e., '*' represents p $\leq$ 0.05, '**' represents p $\leq$ 0.01,' ***' represents p $\leq$ 0.001, ' ****' represents p $\leq$ 0.0001).

The mean value of the usability scores increased in the following order: VR controller (visual) (2.27) < VR controller (audio-visual) (2.82) < Haptic (visual) (4.18) < Haptic (audio-visual) (4.54).

There were statistically significant differences between the following modes: VR controller (visual) vs. VR controller (audio-visual) (p = 0.04 < 0.05); VR controller (visual) vs. haptic (visual) (p < 0.0001); VR controller (visual) vs. haptic (audio-visual) (p < 0.0001); VR controller (audio-visual) vs. haptic (visual) (p = 0.00034 < 0.001); VR controller (audio-visual) vs. haptic (audio-visual) (p < 0.0001); haptic (visual) vs. haptic (audio-visual) (p = 0.04 < 0.05).

The haptic (audio-visual) mode was preferred by most participants, followed by haptic (visual). The next preferred mode was the VR controller in the audio-visual mode. The VR controller in the visual-only mode was least preferred by participants.

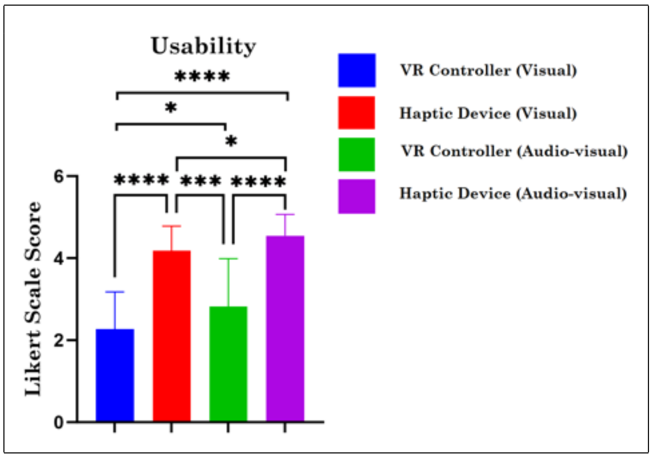

Figure 8: Bar charts comparing evaluation metrics for different modality options. The asterisks in the bar graphs show significant differences between modes. (* represents p $\leq$ 0.05, ** represents p $\leq$ 0.01, *** represents p $\leq$ 0.001, **** represents p $\leq$ 0.0001).

Users preferred the audio-visual modes with both interfaces, demonstrating statistically significant differences from the visual-only modes. For centuries, live music has been offered at many bars, restaurants, and shops to help people relax and cover lulls in conversation. Audio integration inherently lifted users' mood and productivity.

In their research on the emotional response to the value of visualization, Wang et al. [29] discussed how the sense of touch closely connected shoppers with items while shopping and motivated them to purchase those items. Affective haptics [33] generated a sense of touch during remote conversations and established a closer connection between the parties. In the same way, virtual touch played a crucial role in user interaction with the models in our approach. Though users preferred the audiovisual mode with both interfaces, exploring objects with the sense of touch was preferred by most users, demonstrating significant statistical differences with other mode and interface combinations, i.e., VR-controller (visual) and VR-Controller (audio-visual).

Here are some representative comments from participants that complemented the evaluation scores:

"The audio-visual mode transformed the dataset into a musical instrument! I liked to play it with a haptic device."

"The haptic device was very responsive and intuitive."

"I really liked the haptic feedback from the device; it made it easier to tell what I was touching."

"The haptic feedback relates to real life when you touch an object."

"With the controller, it was hard to tell when I was touching the surface. ... I think there should be some visual feedback when the controller is close to the model."

Participants liked the audio integration in visualization. Here are some comments about the audio feedback:

"Audio feedback adds to the experience and understanding of the data."

"Sounds are distinguishable from each other."

"Each color and audio being distinct made it better."

"The audio was good!"

Timbral displays with different musical instruments provided distinct audio feedback to differentiate the minerals from each other and enriched the user experience as they interacted with the data using different interfaces.

6. CONCLUSIONS

In the auditory display, timbres can demonstrate wide variations in the datasets. Virtual touch enriched the visualization experience. In the current version, minerals were arbitrarily located on the model. Future work will incorporate accurate locations of underlying minerals where spatial audio representations will match the spatially located visual representations of minerals. Color brightness and sound cues will be synchronized to demonstrate the ranking of minerals in addition to their presence. Hybrid timbres lying between several musical instruments will be experimented with to demonstrate variations of a mineral within a particular area [34, 35]. In addition, some standard methods for auditory display (i.e., NASA-TLX, BUZZ, and PAMPAS) [36, 37, 38] will be explored.

Participants' feedback will be incorporated into the software design. Some visual cues will be added to the scene as the controller touches a model. The proposed approach will investigate applications such as topological differences, contour maps, and more. The scope of participants will be broadened to get diverse feedback. As part of it, the blind and visually impaired (BVI) will be included in the usability study.

7. ACKNOWLEDGMENT

The authors are thankful to all associated with creating and uploading the following content on YouTube: `https://www.youtube.com/watch?v=VRAXK4QKJ1Q`

8. REFERENCES

[1] S. Jung, A. Wood, S. Hoermann, P. Abhayawardhana, and R. Lindeman, "The Impact of Multi-sensory Stimuli on Confidence Levels for Perceptual-cognitive Tasks in VR," in *Proc. of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Atlanta, GA, USA, March 2020, pp. 463–472.

[2] D. Martin, S. Malpica, D. Gutierrez, B. Masia, and A. Serrano, "Multimodality in VR: A Survey," *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–36, Sep 2022.

[3] C. Donalek, G. Djorgovski, S. Davidoff, A. Cioc, A. Wang, G. Longo, J. Norris, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, and A. Drake, "Immersive and Collaborative Data Visualization Using Virtua Reality Platforms," in *Proc. of the2014 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, October 2014.

[4] M. Batty, "Virtual Geography," *Futures*, vol. 29, no. 4, pp. 337–352, 1997.

[5] S. Ellis, "What are Virtual Environments?" *IEEE Computer Graphics and Applications*, vol. 14, no. 1, pp. 17–22, 1994.

[6] A. Hunt and T. Hermann, "The Importance of Interaction in Sonification," in *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*, Sydney, Australia, July 2004.

[7] N. N. Correia and A. Tanaka, "AVUI: Designing a Toolkit for Audiovisual Interfaces," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Denver, Colorado, USA: Association for Computing Machinery, May 2017, pp. 1093–1104.

[8] J. Middleton, J. Hakulinen, K. Tiitinen, J. Hella, T. Keskinen, P. Huuskonen, J. Culver, J. Linna, M. Turunen, M. Ziat, and R. Raisamo, "Data-to-music Sonification and User Engagement," *Front. Big Data*, vol. 6, pp. 17–22, 2023.

[9] D. Ariansyah, J. A. Erkoyuncu, I. Eimontaite, T. Johnson, A.-M. Oostveen, S. Fletcher, and S. Sharples, "A Head Mounted Augmented Reality Design Practice for Maintenance Assembly: Toward Meeting Perceptual and Cognitive Needs of AR Users," *Applied Ergonomics*, vol. 98, January 2022.

[10] P. R. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, and J. Clow, "Quickset: Multimodal Interaction for Distributed Applications," in *Proceedings of the fifth ACM international conference on Multimedia*, Seattle, WA, USA, November 1997, pp. 31–40.

[11] N. Rönnberg, "Musical Sonification Supports visual Discrimination of Color Intensity," *Behaviour & Information Technology*, vol. 38, pp. 1028–1037, August 2019.

[12] G. Kramer, "Mapping a Single Data Stream to Multiple Auditory Variables: A Subjective Approach to Creating a Compelling Design," in *Proceedings of the Third International Conference on Auditory Display*, Palo Alto, California, November 1996.

[13] B. B. Tomlinson, B. B. Walker, and B. Moore, "Auditory Display in Interactive Science Simulations: Description and Sonification Support Interaction and Enhance Opportunities for Learning," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, April 2020, pp. 1–12.

[14] L. Zhao, L. Zhang, Z. Wu, Y. Chen, H. Dai, X. Yu, Z. Liu, T. Zhang, X. Hu, X. Jiang, X. Li, D. Zhu, D. Shen, and T. Liu, "When Brain-inspired AI meets AGI," *Meta-Radiology*, vol. 1, no. 1, 2023.

[15] E. Burns, S. Razzaque, A. Panter, M. Whitton, M. McCallus, and F. Brooks, "The Hand is Slower than the Eye: A Quantitative Exploration of Visual Dominance over Proprioception," in *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, Bonn, Germany, March 2005, pp. 3–10.

[16] C. Spence, J. Lee, and N. V. der Stoep, "Responding to Sounds from Unseen Locations: Crossmodal Attentional Orienting in Response to Sounds Presented from the Rear," *European Journal of Neuroscience*, vol. 51, October 2017.

[17] N. V. der Stoep, A. Serino, A. Farnè, M. D. Luca, and C. Spence, "Depth: The Forgotten Dimension in Multisensory Research," *Multisensory research*, vol. 29, January 2016.

[18] M. H. W. Rosli and A. Cabrera, "Gestalt Principles in Multimodal Data Representation," *IEEE Computer Graphics and Applications*, vol. 35, no. 2, pp. 80–87, 2015.

[19] I. Rakkolainen, A. Farooq, J. Kangas, J. Hakulinen, J., J. Rantala, M. Turunen, and R. Raisamo, "Technologies for Multimodal Interaction in Extended Reality - A Scoping Review," *Multimodal Technologies and Interaction*, vol. 5, no. 12:81, 2021.

[20] B. Ens, S. Goodwin, A. Prouzeau, F. Anderson, F. Wang, S. Gratzl, Z. Lucarelli, B. Moyle, J. Smiley, and T. Dwyer, "Uplift: A Tangible and Immersive Tabletop System for Casual Collaborative Visual Analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1193–1203, February 2020.

[21] Y. Zhang, B. Ens, K. Satriadi, A. Prouzeau, and S. Goodwin, "Timetables: Embodied Exploration of Immersive Spatio-temporal Data," in *2022 IEEE on Conference Virtual Reality and 3D User Interfaces (VR)*, Christchurch, New Zealand, March 2022, pp. 599–605.

[22] M. I. Billen, O. Kreylos, B. Hamann, M. Jadamec, L. H. Kellogg, O. Staadt, and D. Y. Sumner, "A Geoscience Perspective on Immersive 3D Gridded Data Visualization," *Comput. Geosci.*, vol. 34, no. 9, pp. 1056–1072, Sep 2008.

[23] B. Menelas, M. Ammi, P. Bourdot, and S. Richir, "Survey on Haptic Rendering of Data Sets: Exploration of Scalar and Vector Fields," *Journal of Virtual Reality and Broadcasting*, vol. 6, January 2009.

[24] T. Tarvainen, S. Sapon, and J. Jarva, "Applying Heatmaps in Interpretation of Geochemical Baseline Data on Urban Soils in Finland," *Journal of Geochemical Exploration*, vol. 205, October 2019.

[25] F. Hruby, "The Sound of Being There: Audiovisual Cartography with Immersive Virtual Environments," *KN - Journal of Cartography and Geographic Information*, vol. 69, pp. 19–28, January 2019.

[26] B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 3rd ed. Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1997.

[27] P. C. Wong and J. Thomas, "Visual Analytics," *IEEE Computer Graphics and Applications*, vol. 24, no. 5, pp. 20–21, Sep 2004.

[28] G. Singh, "All That Data Jazz," *IEEE Computer Graphics and Applications*, vol. 44, no. 01, pp. 5–7, Jan 2024.

[29] Y. Wang, A. Segal, R. Klatzky, D. Keefe, P. P. Isenberg, J. Hurtienne, E. Hornecker, T. Dwyer, S. Barrass, and T. Rhyne, "An Emotional Response to the Value of Visualization," *IEEE Computer Graphics and Applications*, vol. 39, no. 5, pp. 8–17, Sep 2019.

[30] R. Wang, C. Jung, and Y. Kim, "Seeing through Sounds: Mapping Auditory Dimensions to Data and Charts for People with Visual Impairments," *Computer Graphics Forum*, vol. 41, no. 3, pp. 71–83, 2022.

[31] S. Barrass, "A Perceptual Framework for the Auditory Display of Scientific Data," *ACM Trans. Appl. Percept.*, vol. 2, no. 4, p. 389–402, oct 2005.

[32] R. Lyon and S. Shamma, *Auditory Representations of Timbre and Pitch.* New York, NY: Springer New York, 1996.

[33] D. Tsetserukou, A. Neviarouskaya, H. Prendinger, N. Kawakami, and S. Tachi, "Affective Haptics in Emotional Communication," in *Proceedings.2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, Amsterdam, Netherlands, September 2009, pp. 1–6.

[34] D. Wessel, "Timbre Space as a Musical Control Structure," *Computer Music Journal*, vol. 3, 1979.

[35] C. L. Krumhansl, "Why is Musical Timbre so Hard to Understand," in *Structure and Perception of Electroacoustic Sound and Music*, S. Nielzén and O. Olsson, Eds. Elsevier, 1989, pp. 43–53.

[36] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds. North-Holland, 1988, vol. 52, pp. 139–183.

[37] T. Ziemer and H. Schultheis, "PAMPAS: A PsychoAcoustical Method for the Perceptual Analysis of Multidimensional Sonification," *Frontiers in Neuroscience*, vol. 16, 2022.

[38] B. J. Tomlinson, B. E. Noah, and B. N. Walker, "BUZZ: An Auditory Interface User Experience Scale," in *CHI EA '18: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, QC, Canada, April 2018, p. 1–6.