A Gesture-Based Approach to Spatialization in Dolby Atmos

Vitor Cansian Sakai Pinheiro
Departamento de Artes Universidade Federal do Paraná
Curitiba, Brazil

Figure 1: Leap Motion Controller being used to control spatialization in Dolby Atmos format.

Abstract

This paper presents a system for the spatialization of sound objects in the Dolby Atmos format, implemented through the integration of an infrared sensor with a chain of three software tools. The setup enables translating hand gestures into spatialization data within the constraints of the Atmos format. The design and parameter mapping are described, along with its usability, strengths, and limitations, as assessed through a preliminary evaluation conducted by the author. Beyond the technical aspects, this article reflects on the author's experience using the system as a mixing engineer and connects these insights to the conceptual framework of related works. This perspective offers a critical reflection on spatialization as a performative practice within studio workflows, highlighting how such devices may be integrated into the multimodal studio environment to introduce new means of interaction in sound spatialization.

Keywords

Spatialization, Dolby Atmos, Leap Motion, Music Mixing, Gesture-based systems

NIME '25, June 24–27, 2025, Canberra, Australia

1 Introduction

Spatialization has become a cornerstone of contemporary music production, propelled by the growing adoption of immersive audio formats such as Ambisonics, Dolby Atmos, and Sony 360. These technologies allow composers and mixers to position sound objects in a three-dimensional space, creating novel auditory experiences for listeners. Over the past two decades, research in the NIME (New Interfaces for Musical Expression) community has laid a strong foundation for innovative interaction design, from early gesture-based systems like data gloves and joysticks to sophisticated motion-capture and multitouch technologies.

Building on this legacy, this paper presents a gesture-based system for controlling spatialization within the Dolby Atmos mixing framework. Inspired by Yulik Yagudin's experiment involving a Leap Motion sensor and the Dolby Atmos Panner in Pro Tools [2], the current research refines those gesture mappings and reflects on their integration into an Atmos workflow from the user perspective. By employing an infrared sensor to capture hand movements in real time, the system provides an intuitive interface for manipulating the spatial properties of sound objects, reimagining the spatialization process in an immersive format used in music, cinema, concerts, and streaming platforms.

Grounded in the concept of space as a structural musical parameter, this research frames the system as a New Interface for Spatial Musical Expression (NISME) [3]. Through practice-based reflections from the author's use, the system's usability, creative potential, and limitations are examined, forming part of a broader

Master's-Degree study on spatialization strategies during the creation and mixing of a song. By enriching the discourse on immersive audio, this research demonstrates how gesture-based systems can expand the expressive scope of spatialization and contribute to emerging workflows in music mixing. To appreciate that contribution fully, we first set it against the lineage of earlier gestural and multichannel approaches.

Long before Dolby Atmos rose to prominence as the commercial flag-bearer for object-based audio, composers and engineers were already performing spatial gestures with Ambisonic rigs, wave-field-synthesis arrays, and multichannel-diffusion systems such as BEAST. To support those practices, a small set of gesture controllers emerged—many first presented to the NIME community. [6] conduct a review of these interfaces and observe that most were one-off prototypes, hindering standardisation and reuse across projects. The present article revisits gesture-based spatialisation within Dolby Atmos—a format backed by a broad industrial standard—in order to bridge that experimental heritage with the practical workflow constraints faced by contemporary mixing engineers.

2 The Dolby Atmos Format

Dolby Atmos, introduced by Dolby Laboratories in 2012, is an immersive audio format that can use up to 118 sound objects with metadata for positioning (x, y, z) and size. These objects are dynamically rendered in accordance with various playback environments, from simple headphone setups to large-format cinemas, adapting the master file for different configurations. Consequently, Atmos is widely integrated in streaming services, theaters, and numerous performance venues.

From a content-creation standpoint, spatialization in Dolby Atmos revolves around panner plugins that control four fundamental parameters: x, y, z coordinates, plus the "size" of each sound object. The size parameter effectively determines the spread of that object across multiple speakers. Typically, positioning can be handled either by manipulating the numerical parameters directly (e.g., adjusting an x-axis slider) or through a 'drag-and-drop' interface, in which objects appear as orbs within the panner window.

When automating object movements over time, two main strategies tend to dominate. First, engineers can create and edit breakpoints in automation lanes for x, y, z and size, drawing curves that define how objects move or change in size throughout the timeline. This approach aligns with established DAW workflows but becomes cumbersome in a 3D setting, where one must often adjust multiple automation lanes to coordinate a single movement. A second method involves real-time mouse capture: as the user plays back the track, they can click and drag objects inside the panner interface, recording their motions to automation lanes. However, translating three-dimensional motion via a two-dimensional mouse interface can limit the fluidity of spatial gestures. It is this challenge—mapping 3-D movements in a more intuitive manner—that motivates the gesture-based system described in this paper.

3 System Design

Atmos only offers four parameters for sound spatialization. Three of them—x, y, and z—are designated "Sound Source Position and Orientation," while "size" is classified as a "Sound Source Characteristic" [4]. Without additional controls (e.g., room properties, directivity, presence/distance, brilliance/warmth) found in other platforms like IRCAM's Spatialisateur or Wave-Field Synthesis, Atmos remains limited. Nonetheless, this narrower scope makes a control device easier to design and implement, simplifying the user's workflow. Additionally, it supports easier adoption in studio environments.

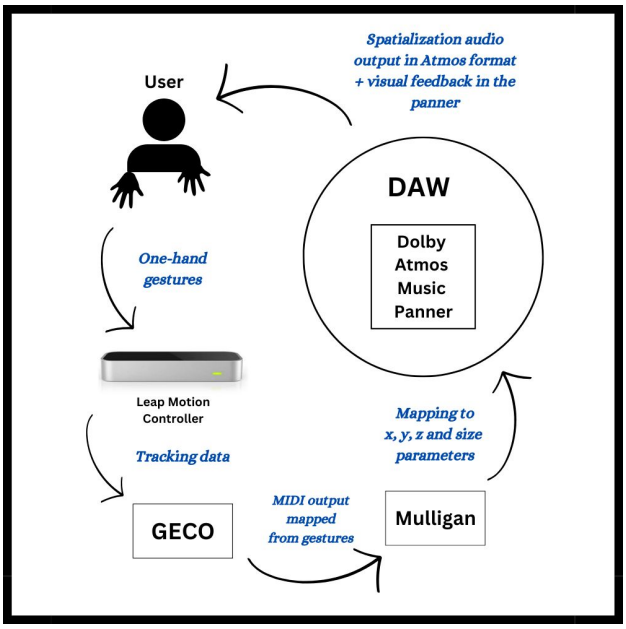

Figure 2: Diagram of the interaction design featuring Leap Motion Controller, GECO and Mulligan Softwares, and DAMP inside the DAW.

The Leap Motion Controller (produced by Ultraleap) was selected as the primary interface. Key factors included its precise hand tracking, low latency, and suitability for a seated position—typical of most professional audio mixing environments. Its established use in the NIME community, combined with low cost and portable design, further supported this choice. Figure 1 shows the Leap Motion positioned upright in front of the console.

To harness the Leap Motion's tracking data, the system uses GECO, software developed by UWYN that translates gestures into MIDI Control Change (CC) or Open Sound Control (OSC) messages. GECO can interpret up to 40 gestures, capturing details such as hand coordinates, hand rotations, and whether the hand is open or closed. This functionality enables users to fine-tune MIDI mappings with parameters for intensity, offset, and resting values. For this study, MIDI was chosen as the primary output protocol.

Since the native Dolby Atmos panner in Pro Tools only supports EUCON control and lacks MIDI functionality, the Dolby Atmos Music Panner (DAMP) plugin is employed instead. Additionally, Mulligan by ReFuse was used, bridging GECO's MIDI-CC messages into the DAMP plugin. This workflow enables control of x, y, z and size parameters through gestures of one hand. A simplified data-flow is summarised in Figure 2.

In terms of setup, the Leap Motion Controller is placed in front of the mixing engineer, facing upward. This orientation aligns its field of view with the user's natural hand movements above the console. Each axis (x, y, z) corresponds to the hand's position, while rotating the hand adjusts the size parameter. A single hand can manage all four parameters concurrently, streamlining the

Figure 3: Dolby Atmos Music Panner plugin.

process of shaping an object's location and spread. Because the user's head is treated as the origin of the virtual field, moving the hand closer to one's body can position the object behind the listening perspective, without turning around. This layout helps maintain an optimal listening position, an approach previously emphasized in gesture-based mixing research [7], while also minimizing ergonomic strain during longer sessions.

4 Preliminary Evaluation

This article documents the preliminary phase of a broader research effort, attempting for empirical considerations to gather initial insights. The author, serving as both designer and user, mixed a song using the Leap Motion-based system and recorded observations on workflow, usability, and creative potential. Although this internal review is not a substitute for large-scale user studies, it provides early feedback that informs subsequent development stages.

A key advantage revealed by this experience was the system's capacity for fluid three-dimensional gestures in real time. Figure 3 illustrates the Dolby Atmos Music Panner interface used during the trial. Movements that once required separate automation curves for azimuth and elevation could now be captured via continuous hand motions. For example, a helical movement around the listener's position might be performed directly, rather than broken down into discrete edits. Additionally, making the "size" parameter accessible through simple hand rotation encouraged more frequent exploration of object spread. Techniques such as synchronizing size pulsations with the musical beat, or dynamically widening a sound source as it rises above the listener, became more intuitive in this gesture-controlled setup.

Despite these benefits, several limitations emerged during these preliminary trials. First, the Leap Motion struggled with rapid or abrupt movements, occasionally requiring post-recording "clean-up" of automation curves. Second, tracking consistency proved vulnerable when hands slipped outside the sensor's field of view, terminating the recorded gesture and prompting restarts. Third, the physical demands of holding arms aloft introduced fatigue over longer periods—after about an hour, the author reported noticeable back and shoulder discomfort. Finally, unintended activations surfaced when a hand casually passed over the sensor, indicating a need for an easily toggled "off" mode or a mechanism to pause gesture capture during routine control-surface activities.

5 Discussion

By integrating gesture-based control in a mixing environment, this system points to potentially more intuitive ways of shaping 3D sound fields. Real-time hand gestures can yield a direct, embodied approach to controlling object positions, circumventing the need to edit or draw multiple automation curves for complex movements. This shift reframes spatialization as a kind of performative act, in line with ongoing discussions in NIME regarding how embodied interaction can facilitate artistic expression. The "gesture zone" created above the sensor, in effect, externalizes the user's mental model of spatial placement, illustrating a form of extended cognition. The user's intent is no longer abstractly manipulated via separate automation lanes but instead expressed physically through real-time gestures, comparable to a conductor guiding sonic elements [4].

However, the absence of tactile feedback remains a common drawback for mid-air interfaces. Haptic cues often enhance performance by reinforcing a sense of physical interaction, which can be especially beneficial when performing intricate or rapid gestures. Some researchers have proposed wearable vibrotactile devices or other forms of local feedback to address these limitations, potentially improving precision and user confidence (see related discussions in [1]). Ergonomically, extended reliance on raised-arm gestures can result in strain or fatigue, suggesting that future prototypes could incorporate smaller, more localized gestures, or alter the sensor's placement to mitigate discomfort.

Equally important is the reliance on multiple layers of software—GECO, Mulligan, and the DAMP plugin—each subject to separate development and updates. While this modular approach offers flexibility, it can complicate troubleshooting and threaten long-term compatibility if any component becomes unsupported. The reliance on multiple software layers introduces potential usability concerns, a challenge also observed in modular musical frameworks like Puara [5]. A consolidated tool, or at least a streamlined integration strategy, would reduce these potential complications. Still, for research and prototyping, the modular architecture has its merits, making it straightforward to replace or upgrade specific components without overhauling the entire system.

With immersive audio gaining prominence in music production, cinema, and VR/AR experiences, developing more intuitive control paradigms is timely. Beyond creative mixing, such interfaces could be applied to live performance scenarios, where

real-time spatial manipulation can enhance audience engagement. The next phase of this research will focus on broader user evaluations with external participants, combining qualitative feedback on ease of use and creative engagement with quantitative metrics on tracking accuracy, gesture recognition, and user fatigue. These studies aim to refine both hardware placement and software mappings, potentially integrating vibrotactile wearables to confirm whether immediate feedback substantially improves performance.

6 Conclusion

This paper introduced a gesture-based system for spatializing sound objects in Dolby Atmos by integrating a Leap Motion Controller with a chain of software tools that transform hand movements into automation data. The findings from this reseach's pilot stage indicate that this setup can streamline the process of executing fluid, multidimensional movements, promoting a more embodied perspective on mixing that treats spatialization as a performative component.

Nevertheless, the results also highlight areas needing further investigation, including limitations related to rapid gestures, sensor field-of-view issues, physical fatigue, and the absence of haptic feedback. Additional development might involve refining ergonomic design, testing vibrotactile or other feedback mechanisms, and simplifying the software chain to enhance reliability. As immersive audio formats continue to expand in entertainment and interactive media, embracing novel interaction models like the one described here may reshape how producers and engineers approach spatial music production. By situating this research at the intersection of NIME, immersive audio technology, and performative mixing techniques, the project underscores the ongoing potential for creative innovation in spatialization controllers in both studio and live contexts.

7 Ethical Standards

This research was conducted in accordance with the highest standards of ethical and professional conduct. The project was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), a foundation within the Ministry of Education in Brazil; Potential non-financial conflicts of interest have been considered and none are deemed relevant to the outcomes of this investigation. All software and tools used in this study have been utilized within the terms of their respective licenses. The methods and processes adhered strictly to ethical guidelines concerning intellectual property and data handling, ensuring transparency and reproducibility in all phases of the research.

Acknowledgments

To Ludiana, Aria, Anina and Aviva, whose love and support carried me throughout the journey. To Clayton Mamedes for his excellent and committed supervision. To the Federal University of Paraná for its institutional support of this research and publication.

References

[1] Anders Eskildsen and Mads Walther-Hansen. 2020. Force Dynamics as a Design Framework for Mid-Air Musical Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Birmingham, UK, 361–366. https://doi.org/10.5281/zenodo.4813418

[2] Production Expert. 2017. Tip: Using Leap Motion Sensor To Control Dolby Atmos Panner In Pro Tools. Web page. https://www.production-expert.com/home-page/2017/10/21/tip-using-leap-motion-sensor-to-control-dolby-atmos-panner-in-pro-tools Accessed: 2025-01-23.

[3] W. Joo, D. Sardana, and I. I. Bukvic. 2020. New Interfaces for Spatial Musical Expression. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'20)*. Birmingham, UK, 21–25.

[4] M. T. Marshall, J. Malloch, and M. M. Wanderley. 2007. Gesture Control of Sound Spatialization for Live Musical Performance. In *Proceedings of the International Gesture Workshop*. Springer-Verlag Berlin Heidelberg, Athens, Greece, 227–238.

[5] E.A.L. Meneses, T. Piquet, and J. Noble. 2023. The Puara Framework: Hiding Complexity and Modularity for Reproducibility and Usability in NIMEs. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. https://nime.org/proceedings/2023/nime2023_11.pdf

[6] Andreas Pysiewicz and Stefan Weinzierl. [n. d.]. Instruments for Spatial Sound Control in Real Time Music Performances. A Review. In *Musical Instruments in the 21st Century* (Singapore, 2016).

[7] J. Ratcliffe. 2014. Hand Motion-Controlled Audio Mixing Interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. Goldsmiths, University of London, UK, 136–139.