

The Immersive Point of Audition: Object-Based Mixing and Domestic Audiovisual Experience

Raphaela Benetello

CITE THIS ARTICLE

Benetello, Raphaela; 2026. The Immersive Point of Audition: Object-Based Mixing and Domestic Audiovisual Experience. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432, doi: 10.18580/setijbe.2026.3 web link: <https://dx.doi.org/10.18580/setijbe.2026.3>



COPYRIGHT This work is made available under a Creative Commons Attribution-NonCommercial 4.0 International License.
Reproduction in whole or in part is permitted provided the source is acknowledged.

The Immersive Point of Audition: Object-Based Mixing and Domestic Audiovisual Experience

Raphaela Benetello

Abstract—This paper presents the concept of immersive point of audition as an original theoretical contribution from doctoral research on object-based audio mixing technologies, namely Dolby Atmos and MPEG-H, and their application to domestic audiovisual consumption. Drawing on cinema sound theory and the established notion of point of audition, the research proposes the immersive point of audition as an auditory experience in which three-dimensional sonic spatiality constructs a perceptual sharing between spectator and diegesis, generating sensations of incorporation and co-presence. A film analysis of *Society of the Snow* (J. A. Bayona, 2023), available on Netflix with Dolby Atmos support, examines how object-based mixing strategies concretize the immersive point of audition across three sequences: the plane crash, the avalanche, and the final crossing of the Andes. The paper also considers the implications of MPEG-H adoption for Brazilian TV 3.0 broadcasting, exploring continuities between immersive audio in streaming and terrestrial television.

Index Terms—object-based audio; Dolby Atmos; MPEG-H; TV 3.0; immersive point of audition; streaming; immersive sound; domestic audiovisual

I. INTRODUCTION

OBJECT-BASED immersive audio technologies have transformed the possibilities for creating and receiving audiovisual sound over the past few decades. A resource that was initially exclusive to movie theaters has progressively expanded to domestic environments, driven by the popularization of streaming platforms and the inclusion of virtualized immersive models in compatible devices. In this context, formats such as Dolby Atmos and MPEG-H enable an aesthetic and perceptual reconfiguration of audiovisual listening that extends well beyond the walls of the cinema.

This expansion raises questions that exceed the technical domain. When immersive sound reaches the domestic listener through headphones connected to a tablet, or through a soundbar in a living room, what kind of perceptual experience is produced? Does the spatialization of sonic objects in a three-dimensional field generate a qualitatively different form of spectatorial engagement, one that existing theoretical frameworks in cinema sound studies have not yet fully addressed?

This paper presents results from doctoral research conducted in the Graduate Program in Arts at the School of Fine Arts of UFMG, focused on the sonic possibilities of object-based mixing technologies in domestic use. From an audiovisual perspective, the research proposes the original concept of immersive point of audition to describe and analyze the perceptual experience produced when three-dimensional object-based audio is consumed in domestic settings. The

concept is applied in a film analysis of *Society of the Snow* (J. A. Bayona, 2023), examining three distinct sequences that illustrate different configurations of immersive listening. The paper also points to the possibilities opened by MPEG-H in Brazilian terrestrial television with the implementation of TV 3.0, drawing continuities between the streaming and broadcasting contexts.

II. OBJECT-BASED MIXING: FROM CINEMA TO THE HOME

The history of spatial audio in cinema is inseparable from the history of attempts to place the listener within the sonic field of the narrative. Multi-channel formats such as stereo, 5.1 and 7.1 established conventions for distributing sound across fixed speaker configurations, anchoring specific sonic elements to specific channels and reproducing the horizontal spatial field with increasing precision. These systems, however, are designed for a fixed speaker layout, and reproduction on any other configuration demands a dedicated remix or an automated downmix that can compromise the original spatial intent.

Object-based mixing represents a paradigm shift, because instead of assigning sounds to fixed channels corresponding to specific loudspeakers, this approach treats each sonic element independently, encoding it with three-dimensional positioning metadata that instructs the reproduction system to render it according to the available configuration. A single mix therefore adapts to different systems, preserving the creative intent of the mixing while maintaining backward compatibility with previous formats.

Dolby Atmos, launched in 2012 for movie theaters and expanded to domestic environments from 2015 onward, has become the most widely adopted standard in this model [8]. The system supports up to 128 sonic objects directed toward up to 64 individual channels (beds) in cinema configurations, allowing sounds to be spatialized across horizontal and vertical axes without assignment to specific channels. A panning tool replaces the traditional channel assignment model and uses positional metadata (X, Y, and Z coordinates) instead of locking sound to specific speakers, giving the mixing continuous spatial control over each sonic object. Differentiated rendering for each type of device, including binaural rendering for headphones, further extends the reach of the immersive experience beyond equipped cinemas [5]. The compatibility of Atmos with professional mixing software such as Pro Tools, Nuendo, and DaVinci Resolve, combined with its export to ADM BWF format for streaming encoding, established it as the dominant standard for immersive audio in streaming platforms.

Raphaela Benetello, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, 30190-002, Brazil. E-mail: raphabenetello@gmail.com

The workflow that connects the professional mixing environment to the domestic listener involves some distinct stages. In Atmos production, the mixing works within a Digital Audio Workstation using the Dolby Atmos Renderer, which monitors the mix in real time across the target speaker configuration and generates a master file encoding both the bed channels and the positional metadata for all sonic objects. This master file is exported in ADM BWF format, a standard developed for the formal description of three-dimensional audio mixes and widely adopted for compatibility across professional environments. For streaming delivery, the ADM BWF master is encoded using Dolby Digital Plus, a codec that reduces the file size for transmission over broadband while preserving the object metadata required for rendering at the receiving device. When a viewer streams a film with Atmos support through a compatible application, the decoding and rendering happen locally on their device, which applies the appropriate rendering algorithm for its own output configuration, full speaker array, soundbar virtualization, or binaural headphone rendering. The listener receives, in effect, a version of the mix that has been customized for their specific playback environment without any intervention by the content producer after the original master was delivered.

A specific dimension of this domestic rendering deserves attention in relation to the analysis that follows. Binaural rendering for headphones relies on Head-Related Transfer Function (HRTF) processing, a set of filters that simulate the acoustic transformations produced by the shape of the human head, ears, and torso as sound arrives from different spatial positions. By applying these filters to each sonic object according to its positional metadata, the renderer creates the perceptual illusion that sounds originate from specific points in three-dimensional space, above, below, in front, behind, or to the sides, even though the signal is delivered through two drivers positioned directly at the ears. The effectiveness of binaural spatialization varies with individual anatomy and listening conditions, and no single HRTF model accounts for all listeners equally. Nevertheless, as the analysis presented in this paper demonstrates, binaural Atmos rendering through consumer headphones produces perceptually significant spatial differentiation that is directly relevant to the construction of the immersive point of audition in domestic settings.

In broadcasting, MPEG-H Audio, developed by the Fraunhofer IIS, combines three-dimensional immersive mixing with advanced interactivity features through metadata. The format supports audio representation in three simultaneous modalities: channel-based, object-based, and scene-based (Higher-Order Ambisonics), enabling sound from above, below, and around the listener with quality rated as “excellent” on the MUSHRA scale in verification tests conducted with 341 expert listeners across nine independent laboratories [9]. Beyond its immersive capabilities, the system allows the viewer to adjust dialogue volume, choose between languages, or activate audio description without significant increase in transmission bitrate, individualizing content consumption. As Murtaza et al. [9] demonstrate, immersive spatialization and interactive personalization are not competing dimensions of the MPEG-H system but complementary ones, given that the same audio objects

that construct three-dimensional spatialization are those that enable individual listener control. In Brazil, MPEG-H was selected in 2021 as the mandatory audio standard for TV 3.0, with regular broadcasts already initiated by networks such as Globo, Rede Amazônica, and TV Cultura, and gradual implementation scheduled from this year onward [7].

III. POINT OF AUDITION: THEORETICAL BACKGROUND

The notion of point of audition entered cinema sound discourse through a practical concern that how to represent, through sound, the spatial and subjective perspective of the characters within a scene. The engineer John L. Cass is credited with one of the earliest uses of the term, in 1930, when he described the sound mix of the period as something that “would be heard by a man with five or six very long ears extending in various directions”, a critique of the prevailing practice of combining sound sources without regard for the listener’s spatial position within the diegesis [1].

Rick Altman’s systematic treatment of the concept identifies point-of-audition sound as characterized by volume, reverb level, and other properties representing “sound as it would be heard from a point within the diegesis, normally by a specific character or characters” [1]. For Altman, this mode of sound constitutes one of the principal mechanisms through which cinema draws the spectator into the narrative as an internal listener, located not in the theater seat but within the story space itself. The spectator is invited, as he puts it, not merely to hear but to identify with someone who hears on their behalf, locating them in “the body of the character who hears for us” [1].

Michel Chion [4] extended this analysis by distinguishing two senses of the point of audition: a spatial sense, referring to the point in the represented space from which the sound appears to originate, and a subjective sense, referring to the character whose auditory perspective the spectator shares. Chion also notes the fundamental asymmetry between image and sound in this regard, since while the camera provides a clearly bounded “symbolic eye,” there is no equivalent “symbolic ear” or “symbolic microphone” in cinema sound, since any given sound may be the product of multiple recording sources combined in post-production [4].

The theoretical elaboration of the point of audition also gave rise to proposals for subcategorization that attempted to address limitations in the original framework. Chion himself later developed the notion of the “twelve ears” of cinema, describing distinct modes of auditory position available in filmic narrative — from the technical ear, a fusion of multiple microphone recordings invisible to any single listener, to the character’s ear, the spectator’s ear, and what he calls the Great Ear, a place where everything can be heard but where “no one is there” [4]. This proliferation of categories reflects the recognized difficulty of assigning a fixed listening position to cinematic sound, since unlike the camera, which provides a clearly bounded directional point of view, sound in cinema is omnidirectional, captured through multiple sources, mixed in post-production, and reproduced through speaker systems that distribute it across the entire space of reception. The result

is that “we cannot answer the question of who is hearing by pointing to the microphone” [4].

This difficulty became more acute as scholars attempted to extend point-of-audition theory to television and new media contexts. Høier [13] described the concept as “a complex phenomenon that is hard to describe,” and proposed a sub-categorization distinguishing observational and active modes of spatial point of audition, and individual and personal modes of subjective point of audition, the latter designating access to “more personal experiences that are not shared with other characters,” such as memories, imagined sounds, or the internalized voices of psychological states. Langkjær [14], in turn, criticized the concept’s dependence on the visual field, arguing that its proximity to the point-of-view shot led to an excessive focus on audiovisual correspondence at the expense of the autonomous specificities of listening. These criticisms converge on a common recognition that existing structures for the point of audition were developed for channel-based, horizontally distributed sound systems, and require revision in the face of three-dimensional object-based spatialization.

More recent scholarship has extended the application of these concepts to spatial audio systems. Vidigal and Novais [12] examine what they call the “envelopment aesthetic” in contemporary cinema, describing cases in which multi-channel and object-based systems allow the spectator to adopt the auditory perspective of specific characters with spatial precision previously unavailable. Their analysis of a scene from *Brave* (2012), first production in Dolby Atmos, illustrates how the system can locate the spectator within the diagonal space between two characters, adopting each character’s perspective in sequence as the scene unfolds. The spatial point of audition, which Chion had identified as theoretically available but practically difficult to realize, becomes technically achievable with object-based spatialization.

IV. THE IMMERSIVE POINT OF AUDITION

The established framework of point-of-audition theory, as described in the preceding section, addresses the spectator’s spatial and subjective positioning within the sonic field of the narrative. However, it does not fully account for a mode of listening that object-based spatialization makes possible, which the spectator is not primarily positioned as the auditory proxy of a specific character, but rather enveloped by an acoustic field that constructs a direct sensory and affective participation in the diegetic space.

The notion of technologically mediated perceptual immersion has been explored in media psychology through constructs such as presence and telepresence. Lombard and Ditton [10] define presence as the perceptual illusion of non-mediation, in which the medium disappears and the user experiences content as if physically inserted within it; Biocca and Delaney [11] describe co-presence as the perception of sharing space with medially represented entities or events. These concepts are perceptual and general in scope, developed to describe experiences in virtual reality environments and interactive multimedia systems. The immersive point of audition is distinguished from this field in that it is specifically anchored in the analytical tradition of cinema sound

studies and in the relationship between listening, diegesis, and narrative corporeality. Rather than describing a property of the reproduction system or an autonomous perceptual state, it designates a mode of spectatorial inscription in the sonic scene mediated by the expressive choices of the mix. Immersion, in this sense, is not a byproduct of technology but the result of aesthetic decisions that technology makes possible.

This research proposes the concept of immersive point of audition, understood as a model of auditory experience in which sonic spatiality constructs a perceptual sharing, generating a sense of incorporation and co-presence. The spectator is enveloped by the acoustic field in a three-dimensional manner, participating in the diegesis as if physically inserted within its space. In this mode of listening, sound acts as a vector of sensation, wherein the audience not only hears as a participant in the diegesis but also feels the sensations present in that space.

The immersive point of audition is further characterized by a shift in the sound function within the audiovisual contract. In the framework proposed by Chion [4], sound and image are bound by a relation of mutual modification, in which “we do not see the same thing when we hear; we do not hear the same thing when we see.” Within this contract, the point of audition operates as a representational device, sound signifies the subjective or spatial perspective of a character, and the spectator adopts that perspective as a listening position. In the immersive point of audition, this representational function is not eliminated but displaced. The sound continues to participate in the audiovisual contract, but its operation shifts from signification to sensation, a displacement that resonates with what Elsaesser and Hagener [6] describe as the cinema of the body, a mode of spectatorial engagement in which the film is experienced as sensory environment to be inhabited. The acoustic field does act directly on the body of the spectator, producing an experience of physical co-presence that precedes or exceeds narrative comprehension.

This sensorial dimension of the immersive point of audition connects to what Carreiro, Opolski, and Meirelles [3] describe as the construction of sonic immersion in cinema, that is, the capacity of sound design, through the articulation of density, timbre, spatialization, and dynamics, to generate an envelopment that positions the spectator not as an external observer of the narrative but as a participant in its physical and affective space. Object-based mixing technologies amplify this capacity by extending it to domestic listening contexts, making the three-dimensional acoustic field available through headphones and consumer-grade devices. This configurations, as the present analysis demonstrates, are capable of producing a qualitatively distinct form of spectatorial engagement even outside the controlled conditions of the cinema.

While Chion’s [4] subjective point of audition anchors the listening experience in the represented perspective of a specific character, where the spectator hears what the character hears, the immersive point of audition does not rely on this character-centered anchoring, since operates through the acoustic envelope of the spectator by the sonic field of the scene itself. A field that may be constructed through the spatial arrangement of ambient sounds, physical impact events, structural sounds,

or the collective sonic texture of a group of characters, rather than through the represented subjectivity of any single one. The experience shifts toward bodily and affective involvement, in which the spatialized acoustic field acts directly on the listener's perceptual apparatus, making the diegetic space physically present rather than merely represented.

This condition is amplified by object-based technologies, which distribute and move sound independently of the fixed axes of traditional channel-based formats, creating acoustic fields that surround the listener in three dimensions even through domestic devices such as headphones or soundbars [3]. The immersive point of audition therefore names a new aesthetic condition enabled by these technologies, which the sonic space of the narrative is not reproduced but inhabited.

V. FILM ANALYSIS: *Society of the Snow* (2023)

The film analysis is employed as a methodological instrument for characterizing the immersive point of audition and for understanding the potential of object-based mixing in domestic audiovisual consumption. The analytical scope is necessarily exemplificative, since a broader range of productions could be examined from the same theoretical framework, and different filmic choices could reveal other dimensions of the concept. Other streaming productions with Dolby Atmos support, such as *All Quiet on the Western Front* (Edward Berger, 2022) and *Okja* (Bong Joon-ho, 2017), offer distinct soundscapes that could be examined through the same framework, potentially revealing other configurations of the immersive point of audition. The present analysis focuses on *Society of the Snow* (J. A. Bayona, 2023), a Netflix original production, because it presents a systematic exploration of three-dimensional spatialization across qualitatively distinct sonic situations within a single film, allowing the concept to be examined across multiple configurations.

The comparative listening was conducted on a Samsung S6 Lite tablet, which allows native enabling and disabling of Atmos processing, with JBL Quantum 200 circumaural headphones, isolating the mixing variable as the determining element of the perceptual experience without variation in the playback device. It is acknowledged that analysis conducted by a single listener on a single domestic device constitutes a limited empirical basis; the results presented here are preliminary in nature and point toward future investigations employing controlled listening protocols and multiple participants.

Society of the Snow [2] is based on the real air disaster of October 1972, when a Uruguayan Air Force plane crashed in a remote section of the Andes mountains, initiating a 72-day survival ordeal. The film, directed by Juan Antonio Bayona, privileges a realist aesthetic and constructs a sound design that contributes decisively to the spectator's immersion. The Dolby Atmos mix builds a sonic landscape reproducing natural elements, such as cutting winds, avalanches, snow underfoot, rarefied silence, that simulates both the geographic vastness and the physical and psychological confinement of the characters. The sound design, by Jorge Adrados, Oriol Tarragó, and Marc Orts, winners of the Goya Award, dedicates particular attention to ambient textures, like the fuselage,

breathing, footsteps in snow. The score by Michael Giacchino adopts a discreet and emotional presence, balancing silence with atonal and epic structures, using candombe drums and Mapuche choral voices in scenes of high dramatic intensity, and shifting to open harmonies with guitars and female voices in the film's resolution.

A. *The Plane Crash*

The crash sequence, occurring approximately thirteen minutes into the film, inaugurates the central tragedy of the narrative. Before the aircraft enters the turbulence zone, the sound design establishes a fluid and low-density texture, with dispersed voices, the continuous hum of engines, low-frequency vibrations typical of a pressurized cabin. As the plane enters the Andean turbulence, the sonic atmosphere transforms. Structural vibrations, sudden motor variations, and atmospheric noises move to the foreground, thickening the texture. In the Dolby Atmos version, the tridimensionality of these sounds is immediately apparent. The creaking of the structure arrives from the lateral channels, the vibration of the seats is rendered with bass frequencies positioned in the lower field, and alerts and screams travel across the sonic space from multiple positions. The soundscape fractures into simultaneous layers of short-duration, high-intensity acoustic events, with chairs scraping, structural ruptures, objects piercing the cabin, baggage colliding with the interior walls, culminating in the collision with the mountain. The impact breaks the fuselage, bodies are thrown, screams overlap, and wind infiltrates the sonic space. The sequence ends with the fuselage sliding across the snow to a stop, followed by a fade out in which sound ceases for several seconds. This brief void, following the sonic overload, heightens contrast and repositions the spectator's immersive point of audition in alignment with the subjective state of the survivors, such as disorientation, shock, and the suspension of time.

In the stereo version, these elements remain present but with reduced density and positional precision. The absence of height channels makes the sound distribution more compact, reducing the sense of event dispersion. Sounds tend to overlap in a centralized, homogeneous frontal field, particularly in passages of high complexity, where the distinction between metallic fuselage sounds and the sharp creaks of structural ruptures becomes more difficult to perceive. The stereo version preserves the narrative and emotional coherence of the sequence, but loses part of its spatial expressiveness and the depth of the sensory envelopment that the immersive point of audition describes.

In the moments immediately following the crash, the shocked state of the survivors is rendered through sparse, minimal sonic textures, like breathing, groans, and intermittent structural sounds from the damaged fuselage. This auditory emptying evokes temporal suspension and reinforces the perception of trauma. As the characters regain consciousness, the spatialization of sounds reconstructs the closed, fragmented environment of the destroyed fuselage. In the Atmos mix, the alternation between moments of higher sound amplitude and passages of low sonic intensity constructs a dramatic auditory

progression that positions the spectator inside the cabin rather than before it.

The sound construction of the crash sequence drew on recordings made inside aircraft prototypes and physical materials in contact with snow and metal, lending acoustic verisimilitude to the sequence. The combination of deep bass frequencies with intermittent high-frequency transients accentuates the sense of vertigo and structural disintegration. In the Atmos version, the spatial organization of sounds in the immediate post-crash silence gives the spectator an auditory map of the survivors' positions within the destroyed fuselage before they become visible. Groans and labored breathing arrive from distinct points in the three-dimensional field, metal plates shift and compress against snow from lateral and lower registers, and voices in varying states of distress and confusion create a spatial texture in which each sound source is individually locatable. The alternation between moments of higher amplitude and passages of near-silence constructs a dramatic auditory progression that functions, perceptually, as an extension of the physical space of the cabin, where the spectator inhabits it rather than observing it. The analysis of this sequence thus confirms that the immersive point of audition is not confined to scenes of maximum sonic density, acting equally in the sparse, fragmented sonic texture of shock and disorientation, where the rarefaction of the acoustic field becomes itself a form of corporeal presence.

B. The Avalanche

The avalanche sequence concentrates three dimensions of the immersive point of audition: spatial anticipation, physical envelopment, and the internal listening perspective of the characters.

Before the snow displacement, the survivors are awake inside the fuselage partially buried by accumulated ice in the days following the crash, engaged in a game of verse recitation and rhyming humor. The soundtrack is entirely diegetic, with light density and a predominance of mid-range frequencies, where vocal sounds and discrete laughter construct an intimate ambience that highlights the rarity of the moment. The careful, naturalistic rendering of this brief levity serves as a deliberate sonic contrast for what follows. After a few exchanges, all the survivors laugh and applaud a failed rhyme attempt. During the applause, a low-frequency, progressive sound, initially indistinct, with dense texture and opaque timbre, begins to emerge in the background, gaining body and amplitude, displacing itself in the sonic spectrum before any of the characters identify it, while a sense of fuselage trembling modifies the entire ambience of the scene.

In Dolby Atmos, this sound is spatialized with verticalization and depth, traveling through the upper and rear channels in a way that suggests the massive displacement of snow descending from the top of the mountain. The spectator's immersive point of audition is installed inside the fuselage, sharing with the characters the same delay in perceiving the threat. Sonic density increases abruptly as the avalanche hits with metallic sounds against ice, the wall of luggage protecting against the cold being thrown against the characters,

muffled screams, and above all the impact of snow burying the survivors and nearly all the wreckage. Though this portion lasts less than ten seconds, it produces a perceptible change of rhythm and a flooding of the diegetic plane with elements that are difficult to distinguish individually but that communicate the disorientation and emotional collapse of the moment. In the stereo version, the fundamental elements of the scene are preserved, but the sonic field is less spatially articulated. The avalanche displacement is perceived more as a growing sonic wall than as a complex spatial movement. The collapse is audible but reduced in its sense of physically enveloping the spectator.

After the invasion of snow, a ten-second total fade out in black simulates burial, during which sound ceases entirely. This sonic and visual void reinforces the abrupt interruption of life and amplifies the emotional effect. In the final seconds of the dark screen, muffled sounds emerge suggesting internal digging. The creaking of the fuselage under the pressure of snow, particles falling, and above all the breathing of the survivors, in different intensities, timbres, and rhythms, some rapid and panting, indicating panic or physical effort; others slow, nearly inaudible, reflecting exhaustion or shock. When a character manages to break free, their vocalization sounds distant and filtered. In the Atmos version, this sound is reproduced as if coming from below or behind, simulating the immersive point of audition of the other survivors still buried and their desperate attempts to emerge from the displaced snow mass. Small noises, such as hands removing snow or utensils accidentally shifting, gain relevance filling the rarefied sonic space. In the stereo version, these elements are present but more centralized, the ambience that simulates the weight of snow above the fuselage and the distance between bodies reduced to a flatter sonic layer, technically competent but spatially less precise.

The sequence continues with the survivors digging in a dark, enclosed space, sounds of low density and opaque timbres constructing a sense of confinement, until a second avalanche interrupts the collective effort. A shorter fade out follows, and the return of the visual field reveals that an air pocket has formed, keeping the survivors together. In the Atmos version, the voices of the characters present spatial differentiation in depth and height, allowing a symbolic perception of each person's position in the scenic space and intensifying the sense of confusion and urgency. In the stereo version, sounds tend to cluster, reducing the spatial distinction between elements. There is no musical score until the end of the sequence, which closes with the return of the central character's narration. The construction relies on intermittent silences, low density, and diegetic elements to transmit the physical and emotional impact of burial. The immersive point of audition operates here at its most immediate: more than hearing as the characters hear, the spectator becomes, perceptually, another survivor, experiencing the anguish of frozen burial through muffled sounds, the cries of colleagues, and the sensation of attempting to excavate with hands that cannot move freely.

C. The Final Crossing

The sequence depicting the crossing of the Andes toward Chile, initially undertaken by Fernando Parrado, Roberto Canessa, and Antonio Vizintín, marks a sonic displacement from the intensity of the earlier sequences. Rather than amplifying the brutality of the environment through density and amplitude, the sound design here invests in rarefaction as a narrative strategy. Vocal silence predominates, and ambient sounds, like the wind, the creak of snow underfoot, breathing, take the foreground.

The extradiegetic soundtrack, characterized by organic timbres with ethnic resonances, including Candombe drums and Mapuche vocalizations, structures the open and wide-angle shots of the mountainous landscape, while diegetic sounds, such as footsteps, breathing, and brief dialogues, gain prominence in closer framing. When the climbers reach the summit only to discover an expanse of further mountains ahead, the score ceases and the mountain ambience returns, wind rising in frequency and acuity, accentuating the aridity of the location and the difficulty of the journey.

In the Atmos version, the wind is a dynamic sonic body that moves laterally, horizontally, and at times appears to arrive from above, amplifying the sense of exposure and the vulnerability of bodies before the environment. The precision of this lateral and vertical displacement gives the natural elements a physical presence that the stereo version, in which directionality is limited to horizontal channels, cannot reproduce with equivalent specificity.

This configuration of the immersive point of audition represents a distinct aesthetic use of object-based spatialization. In the crash and avalanche sequences, the three-dimensional field is filled with sounds that arrive from multiple simultaneous directions, from above, below, and laterally, producing an envelopment that is physically overwhelming. In the final crossing, the same three-dimensional field is largely emptied, and what remains, the lateral movement of wind, the localized crunch of snow under specific footfalls and the directional sound of exhausted breathing, carries disproportionate narrative and perceptual weight precisely because the surrounding acoustic space is so sparse. The immersive point of audition here acts through the body's heightened sensitivity to sonic detail in conditions of near-silence, in which sounds that in a denser mix would be masked or peripheral become, in this rarefied field, present. The Atmos rendering makes this heightened presence spatially specific in a way that the stereo version, with its more compressed horizontal field, cannot replicate with equivalent precision. In the stereo version, the wind and footstep sounds are present and effective, but their positional information is reduced to a left-right distribution that does not capture the vertical and depth dimensions through which the Atmos mix conveys the physical exposure of the climbers to the open Andean environment.

As the landscape changes, losing its whiteness, revealing the earth tones of lower altitude, the score grows in presence and amplitude, a harmonic opening that signals narrative resolution. In the Atmos version, this transition is reinforced by an expansion in the auditory field, with greater perception

of height and depth as the wind layers with the musical score. The final sequence, in which the survivors encounter a Chilean farmer, is the first moment in the film in which the sonic field opens fully outward, a spatial liberation that the tridimensional mix renders with a perceptual clarity the stereo version cannot replicate in the same degree.

The three sequences examined here do not exhaust the sonic richness of *Society of the Snow*. Other moments in the film equally merit analysis within the framework of the immersive point of audition. The interior of the fuselage, whose curved and closed metallic structure produces reverb and muffling that the Atmos mix renders with spatial circularity, functions throughout the film as what the sound design team described as a resonance box for the tragedy, where each sound that occurs within it is amplified and reflected by the structure in ways that reinforce the physical and psychological confinement of the survivors. The rescue sequence, in which the survivors are lifted by helicopter and the closing of the aircraft door is followed by a sudden interruption of the ambient sound, is another case in which silence operates as a narrative and spatial event, signaling the end of the Andean ordeal and a perceptual break that the immersive point of audition renders as a form of bodily transition.

VI. DISCUSSION

The three sequences analyzed across *Society of the Snow* illustrate that the immersive point of audition is a listening mode that assumes distinct configurations in response to the expressive demands of each moment. In the crash sequence, it acts through physical envelopment and the simulation of spatial disorientation; in the avalanche, through progressive spatial anticipation and the collective experience of burial; in the final crossing, for the rarefaction and the amplification of individual sounds against an overwhelming silence. In each case, the Atmos version seems to alter the qualitative nature of the spectatorial experience in comparison to the stereo version, not at the level of narrative comprehension, which both versions support effectively, but at the level of perceptual and affective engagement with the sonic space.

This distinction maps onto the broader theoretical claim of the immersive point of audition, that object-based spatialization constitutes a different mode of spectatorial inscription in the sonic field, one that exceeds the categories of subjective and spatial point of audition as theorized by Chion [4] and Altman [1], and that cannot be described adequately through the general presence-theory frameworks of media psychology [10], [11]. The aesthetic decisions encoded in the mix, like the choice to verticalize the approaching avalanche, to position a survivor's cry as arriving from below and behind, to spatialize the wind as a dynamic body moving laterally around the listener, are the operative elements of the immersive point of audition. Technology enables them and the mixer's expressive choices determine whether and how they are realized.

The analysis also raises a methodological question with implications for future research: what instruments are appropriate for the study of the immersive point of audition in domestic contexts? The comparative listening protocol employed here,

that is, same device, same scene, Atmos processing enabled and disabled in alternation, isolates the mixing variable with reasonable precision for a qualitative, single-listener study. It does not, however, account for variability in individual listening sensitivity, familiarity with immersive audio, or the acoustic properties of the listening environment. Future studies might productively combine this kind of perceptual description with spectator response data collected through controlled experiments, extending the empirical base of the concept beyond the interpretive account of a single analyst. The theoretical framework proposed here is intended as a contribution to the vocabulary available for such studies.

The three sequences analyzed from *Society of the Snow* also suggest that the immersive point of audition does not map neatly onto any single technical parameter of the Atmos mix. It is not produced by height channels alone, nor by object positioning alone, nor by the binaural rendering applied to headphone output. It emerges from the interaction between these technical resources and the expressive decisions of the sound designer, which elements to spatialize and in which direction, how density and silence are alternated, what timbral qualities are assigned to ambient sounds, and how spatial movement is coordinated with the narrative arc of the scene. This means that the immersive point of audition is an aesthetic category as much as a perceptual one, because it names something that the mix does, not something that the technology does automatically.

The implications for the broadcasting context are also significant. The analysis demonstrates that the immersive point of audition is achievable in domestic listening conditions through headphones, one of the most accessible and widespread playback configurations available to contemporary audiences. This matters for the broadcasting context because it means that the perceptual experience documented here does not depend on specialized equipment, it is already available to anyone with a compatible streaming subscription and a pair of standard headphones. MPEG-H, in the Brazilian TV 3.0 standard, also supports the three-dimensional object-based rendering that makes this experience possible, combining it with interactive personalization features that extend the listener's agency over the sonic field. As Murtaza et al. [9] confirm, the objects that produce spatial envelopment are the same objects that can be individually adjusted or activated by the viewer, meaning that immersivity and personalization are produced by the same underlying architecture.

The transition from streaming to terrestrial broadcasting that TV 3.0 represents therefore carries a specific implication for the concept proposed here. In the streaming context, the immersive point of audition is a conditional experience that requires a compatible subscription tier, a device with Atmos decoding, and a content producer who has invested in an immersive mix. In the TV 3.0 context, with MPEG-H as the mandatory audio standard, these conditions are structurally embedded in the broadcast signal itself. Any compatible receiver, including the set-top boxes and adapted televisions that will accompany the gradual rollout of the new standard across Brazilian territory, will have native capacity to render object-based spatial audio. The democratization this represents

opens the possibility that the kind of perceptual and affective engagement described here as the immersive point of audition becomes a constituent feature of everyday domestic television viewing, rather than a premium experience available only to a subset of streaming audiences.

VII. FINAL REMARKS

The research suggests that object-based mixing technologies play a role in the aesthetic reconfiguration of audiovisual listening that extends beyond technical innovation into the domain of perceptual and narrative experience. The concept of immersive point of audition, proposed as an original theoretical contribution, offers an analytical instrument for understanding how immersive sound, even when reproduced on domestic devices, is capable of displacing the spectator's experience from a position of observation to one of presence and corporeality within the scene.

The analysis of *Society of the Snow* illustrates qualitative differences between Atmos and stereo versions across three distinct sequences, suggesting that the immersive point of audition assumes multiple configurations depending on the expressive strategies of the mix and the dramatic demands of each moment. These results are preliminary and subject to the methodological constraints of a single-case, single-listener study. Future research with controlled listening protocols and multiple participants would be necessary to substantiate the perceptual claims advanced here with broader empirical grounding.

With the progressive implementation of TV 3.0 in Brazil, the possibilities for immersive listening and personalized audio consumption are set to expand to terrestrial television, democratizing access that currently still depends on streaming platforms and compatible devices. This development makes it relevant to foster creative and analytical practices that consider the specificities of spatial sound across the different exhibition windows of contemporary audiovisual media, and to develop theoretical frameworks, such as the one proposed here, that can account for what changes, perceptually and aesthetically, when the listener is no longer before the sound field and instead feels immersed in it.



Raphaela Benetello received the Ph.D. degree in Arts, with an emphasis on Cinema, from the Graduate Program in Arts at the Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil. Since 2017, she has been an Audiovisual Technician at UFMG. Her research interests include audiovisual sound, immersive audio technologies, object-based mixing, and domestic audiovisual consumption.

REFERENCES

- [1] R. Altman, "Sound Space," in *Sound Theory/Sound Practice*, R. Altman, Ed. New York, NY, USA: Routledge, 1992, pp. 46–64.
- [2] J. A. Bayona, dir., *Society of the Snow* [Film]. Spain: Netflix, 2023.
- [3] R. Carreiro, D. Opolski, and R. Meirelles, *A imersão sonora no cinema*. Salvador, Brazil: EDUFBA, 2023.
- [4] M. Chion, *Audio-Vision: Sound on Screen*. Lisbon, Portugal: Texto & Grafia, 2011.
- [5] Dolby Laboratories, *Dolby Atmos Home Theater Installation Guidelines*. San Francisco, CA, USA: Dolby Laboratories, 2016.
- [6] T. Elsaesser and M. Hagener, *Film Theory: An Introduction through the Senses*. Campinas, Brazil: Papirus, 2018.
- [7] Fraunhofer IIS, *MPEG-H Audio: The Next Generation Audio Codec*. Erlangen, Germany: Fraunhofer IIS, 2023.
- [8] G. Sergi, "Knocking at the door of cinematic artifice: Dolby Atmos, challenges and opportunities," *New Soundtrack*, vol. 3, no. 2, pp. 107–121, 2013.
- [9] A. Murtaza, S. Meltzer, Y. Grewe, N. Faecks, M. Raulet, and L. Gregory, "MPEG-H Audio System for SBTVD TV 3.0 Call for Proposals," *SET International Journal of Broadcast Engineering*, vol. 7, 2021, doi: 10.18580/setijbe.2021.3.
- [10] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, 1997.
- [11] F. Biocca and B. Delaney, "Immersive virtual reality technology," in *Communication in the Age of Virtual Reality*, F. Biocca and M. Levy, Eds. Hillsdale, NJ, USA: Lawrence Erlbaum, 1995, pp. 57–124.
- [12] L. Vidigal and M. Novais, "Ponto de escuta e tridimensionalidade sonora na ficção e no documentário: Para uma estética do envelopamento no cinema contemporâneo," *Revista Tríade*, vol. 8, no. 17, 2020.
- [13] S. Høier, "The Relevance of Point of Audition in Television Sound: Rethinking a Problematic Term," *Journal of Scandinavian Cinema*, vol. 3, no. 1, pp. 37–51, 2013.
- [14] B. Langkjær, *Den lyttende tilskuer: Perception af lyd og musik i film*. Copenhagen, Denmark: Museum Tusculanum, 2000.