Turn Your Head and Listen: 360° Audio Between Old Utopias and Market Strategies

Raffaele Pavoni*

University of Florence (Italy)

Abstract

Many scientific publications directly concern VR, often conceived as a battlefield for rethinking our relationship with the moving image, and our frameworks on topics such as the cinematographic language or the spectator perception. And yet, although for some time film studies have re-evaluated the role of sound in cinematic production and spectatorship, in the field of VR the theoretical reflection on concepts such as stereoscopic audio and binaural recording has remained confined to the technological-engineering aspect. This essay aims to explore the most common audio techniques and their real impact in terms of consumption and affordances, on the basis of some software studies and sound studies reflections and within a media-archaeological perspective. The results of these innovations go beyond the technological features, suggesting a further rethinking of the cinematographic form and calling into question, even from an audio-only perspective, the issue of audience perception and interaction. The case studies taken into account might be an impulse, heuristically, for new studies on these topics.

Keywords: Spatial Audio, Virtual Reality, Binaural Audio, Sound Studies, Software Studies.

* ✉ raf.pa53@gmail.com

1   Introduction

Many recent scientific publications directly concern Virtual Reality (VR). This comes by one hand from the growing availability of products and devices; by the other, from the impulse of revitalising film studies through a medium that appears to be, somehow, destabilizing. VR, in other words, is often conceived as a battlefield for rethinking our relationship with the moving image, and our frameworks on topics such as the cinematographic language or the spectator experience.

And yet, although for some time film studies have re-evaluated the role of sound in cinematic production and spectatorship (often intertwining the paradigms of sound studies, or underlining the aural feature of the cinematic experience), in the specific field of VR studies the theoretical reflection on concepts such as spatial audio and binaural recording has remained confined to the technological-engineering domain. This one seems to be actually a lack of these evolutions and mutations' interpretations.

This article explores some of the most common audio techniques and their impact in terms of consumption and affordances, on the basis of the abovementioned reflections and within a media-archaeological perspective. I will focus on some case studies, both on the hardware and on the software side, which seem to be particularly emblematic: indeed, the results of these innovations go beyond the simple technological fact, suggesting a further rethinking of the cinematographic form.

The utopia of sound immersion is, in fact, consubstantial with the history of cinema, although often relaunched (and withdrawn) for purely commercial reasons. And yet, as this essay will theorise, the abandonment of a Point of Listening (POL), as the abandonment of a Point of View (POV), seems to be at the same time a strength and a limit of those products, calling into question, even from an audio-only perspective, the issue of audience perception and interaction. This kind of approach, built on a software and sound studies methodology, could represent an impulse, heuristically, for new studies on the subject. In the next chapter, I am going to insert this hypothesis in the pre-existing debate on VR, drawing a theoretical framework and applying it to specific technologies, in order to prompt — in the conclusions — further studies on these topics.

2   VR: The State of the Art

A systematisation of VR studies implies a theoretical and academic distinction between a social or aesthetical approach and an engineering one. This phenomenon could be inserted in the broader scenario of an evolution of visual culture, which seems to face a growing predominance of the device, intended as a cultural artefact. At the same time, studies on the technological features of these devices often confine with advertisings (both hiddenly or explicitly). To discern market tactics and academic studies is sometimes as uneasy task, especially in the field of VR, as the emphasis that leads both kind of publications is somehow similar.

Studies, highlighting the educational potential of VR, are increasingly flourishing, especially in a moment where the COVID-19 pandemic has forced the exploration of new teaching strategies. Along this line, for example, we find publications that underline the lack of clear guidance in the library community on how to introduce these technologies in effective ways and make them sustainable within different types of institutions (Cook et al. 2019) or hypothesize a theoretical framework for the development of educational Augmented Reality Audio Games (ARAG) (Rovithis et al., 2019a), or investigate the benefits of VR on music education (Fletcher et al. 2019). Particularly relevant, in this sense, are the contributions of Andreoli (2018), who proposes some significant case studies in terms of gamification of learning, and Rossi (2020), who supports and integrates the knowledge of IT tools for the design and management of virtual environments with indications relating to the organizational and strategic activities that govern the formation of the image in VR.

Another vivid field of studies, related to VR, concerns cultural and digital heritage, focusing on the role of VR in preserving and divulging cultural artifacts. This may lead to establish different connections between architecture, visual arts and education (Luigini and Panciroli, 2018), or between 3D modeling and rendering and the use of documentary materials (Basso 2020). This may also suggest a reconceptualisation of museums in the context of a new exhibition design, paying attention to the inversely proportional relationship between attractiveness and credibility (Amoruso 2019). In the most extreme outcomes, this can lead to a renegotiation

of the very notion of heritage, and of its constituting elements: "are the objects themselves? Is the site as a whole? Is the excavation a documentation?" (March 2001: 277, my translation). Battini (2017) analyses the virtual simulation of architectural sites, whereas Nau (2019), on an archaeological level, questions how to make VR suitable for the analysis of a broader processing spectrum of hypotheses, possibilities and solutions, both in research and in restoration. Moreover, some psychological essays focus on the the effects of VR perception on associative memory (van Helvoort et al. 2020), or on anxiety states and heartbeats (Brinkman et al. 2015).

Finally, as is the case for this essay, VR is a privileged object of film and game studies, although many of these contributions are often fragmented in the multiple streams of individual case studies, however relevant they may be (Jung et al. 2020; Steincke 2016; Arcagni 2018, 2020; Dalpozzo, Negri and Novaga 2018). The lowest common denominator of all these contributions seems to be the desire to conceive virtual reality as an experimental form, where filming and gaming habits intertwine with engineering features, and Humanities are confronted with a forced interdisciplinarity. Many contributions deal with the remediation and reorganization of perception; in particular, VR is often conceived as a disruptive epistemological tool, which entails an emancipatory drive (as in the largely autobiographical work of Lanier) (2017). Phenomenological, neuroaesthetical and ontological approaches are often privileged, in these studies, as they help to rediscuss concepts such as the point of view, the immersiveness of contemporary media, the role of screens in the current media ecology, the perception of virtual experience as a physical on (and vice versa), the identification process and engagement in VR stories, the features of wearable technologies; last but not least, as in this essay, the POL. Despite similarities, as Majkut argues, "the epistemologies of POL and POV are dissimilar in fundamental ways," primarily because ubiquitous POL "isn't 'point' as 'center' of sense orientation. Metaphorically, it may be better thought of as a 'corner of orientation,' located peripherally on a horizon" (2014: 179).

This issue, in particular, has mainly been the subject of engineering analysis, whereas few contributions, in Humanities, tend to highlight the potential of spatial sonic expression. These studies deal with the sound not only as an integral part of immersive virtual experiences, but also as a critical point of departure for creative and technological works (Çamcı and Hamilton 2020), or propose a simple and efficient system to estimate room acoustic for plausible reproduction of spatial audio using 360° cameras for VR / AR applications, based on the estimated geometric and acoustic properties in the scene (Kim et al. 2019). Engineering studies, on their side, deal with concepts such as the spatial separation of binaurally recorded speakers (Kern and Ellermeier 2020) and the telecommunication systems that support them (Gamper and Lokki 2010), or sound interaction in game playing (Rovithis et al. 2019b).

Humanities, to sum up, often seem to leave those topics to the engineering studies. Yet, the contribution of software studies should lead us to a greater sensitivity in this aspect. Specifically, in VR the close interdependence between the medium (the interactive video) and the content (the graphic interface), in which one finds its *raison d'être* only through the existence of the other, suggests an overcoming of both these concepts, focusing on the dialectic between hardware and software. The concept of medium as software, in particular, seem to be useful: considering everything that stands between sender and recipient as a manipulation through software, it is possible to enclose the totality of the communication processes put in place and to analyse them in a unitary framework that takes into account the specificities of the single product. When J. L.Godard ironically stated that "camera movements are a question of morality" (Rivette 1961, my translation), he tended underlining how the "non-neutrality" of the gaze was added to the "non-neutrality" of the support. This support, following Manovich's insight, has its own status: "a code may also provide its own model of the world, its own logical system, or ideology; [...] most modern cultural theories rely on these notions, which together I will refer to the 'non-transparency of the code' idea" (2001: 64).

The question of audio in VR, if analysed from a cultural point of view, brings out the role of social interactions in the definition of a medium, regardless of its commercial success (a failure is, in itself, "social"). The innovation, in fact, transcends the physical support to manifest itself, as I have already said above, in the social use that is made of such supports. This is what Carlo and Colombo argue, for instance, underlining precisely that the main feature that digitization offers to the media is that of detaching them from their supports. After a long-term definition of a medium "on the basis of the connection between a certain technology, a certain language and certain conditions of use" (Carlo and Colombo 2007: 16, my translation), today, inevitably, the definition of medium refers only to some kind of social use. The proposal, the two scholars conclude, is to "consider a single

medium, in a given historical period, as a momentary balance between a multiplicity of social dimensions, which go beyond the medium itself, but also shape it and are in turn lively" (17, my translation). These social dimensions seem to be best expressed by the algorithmic architecture of the medium. Quoting Manovich again, "for users who only interact with media content through application software, the 'properties' of digital media are defined by the particular software as opposed to solely being contained in the actual content" (Manovich 2013: 152). Many studies have followed this path, as Galloway's reflections on the protocol (Galloway 2004) and the interface (Galloway 2012), and those of Chun on the relationship between new media and habit linked to old media (Chun 2016). This framework reveals to be useful, as we will see, in the analysis of VR audio.

3   VR: Where Sound and Software Studies Converge

The attention, in the contributions cited, is to the role of software in defining the aesthetics and interaction methods of contemporaneity. Software defines the visual culture in the double sense of a social shaping of technology and, we could say, of a technological shaping of society, as technology and society are both expressions of a temporally and geographically situated culture. To this regard, by updating and developing Gibson's studies (1979), the affordances of the algorithm and its impact on the filmic experience seem to have both a technological and a social matrix, where both aspects contribute to defining contemporary aesthetics. This is particularly true in the case of VR: to examine the affordances that differentiate these technologies from physical reality may be a valid approach to understanding why users adopt these technologies, wondering, therefore, "whether the historical floundering of virtual and augmented applications has partly been due to a failure to understand the natural affordances of these technologies" (Steffen et al. 2019: 723).

The problem seems to be reconfigured in the environmental and 'ecological' conception of the screen proposed by screen studies, in which "a screen is always part of a screenscape [...] aimed at offering a mediation with the world and the others thanks to both the images that it hosts and its very nature of site" (Casetti 2019: 46). The device is here reconceptualised as an even more interactive interface: In this sense, VR is both an object of study and a heuristic tool, which in the final chapter will be developed as such. Software studies, to this extent, seem to share the same, so to speak, "culturological" matrix, with another line of study that is particularly relevant to our object of study, and on which it is perhaps necessary to dwell for questions of method; that of the so-called sound studies.

Many scholars have headed in this direction, often systematically (Sterne 2012). Frequent is the reference to the notion of "relational space," coined by Scott McQuire to emphasize the role of the media in changing the contours of everyday experience and social space, making it variable and contingent (McQuire 2008); this relational aspect of sound, and sound media, is already a first important sign of contact with the software perspective, suggesting how both perspectives may converge. We may retrace the same convergence also in the reference to Henri Lefebvre (1974), who from a constructivist point of view coins the notion of "lived space," which is experienced through the images and symbols associated with it, or to the studies on soundscapes by Robert Murray Schafer (1977), whose book *The Tuning of the World* is almost a kind of manifesto in this field. Sound studies, like software studies, re-evaluate the device in a cultural and relational sense (as recently film studies tend to do, as well).

Spatial audio debate shares the same theoretical issues. Moving on to our case study, indeed, Parisi shows how the effectiveness of virtual experience depends on how the constraints and possibilities of human bodies are respected: the higher virtual and physical experience are tuned, the higher is the illusion of presence in the virtual world. As the author writes, "thanks to the media we are able to transcend the limits imposed by our biology. We do this every day when we wear glasses to see better; when we see the surface of the moon thanks to the use of powerful telescopes; when we deal with images as if they were real alternative worlds" (Parisi 2020: 90, my translation). Bodily syntonisation stresses the importance of the acoustic side of virtual experience, against the mere visual one. Frameless vision, in fact, seems to be phenomenologically less important than actively exploring the environment by using the body as an interface. To quote Parisi again

> the current limitations to the transparency of virtual reality are attributable to technical and design limits which [...] could be reduced simply by considering the phenomenological characteristics that characterize the embodied experience. Since the body is the first mediator, we cannot

conceive of a virtual experience if not starting from what we are used to in normal ecological behaviour (97, my translation).

Total transparency, the author concludes, "appears impossible, but I would also add useless" (97, my translation). This transparency therefore seems to bring out the device, and this is the contradiction and the one of the main theoretical interests of VR. Here, moreover, VR reveals to be strictly connected to the Early Cinema, that is, to a certain type of cinema of attractions that re-evaluates the device as the very object of the filmic experience. In this sense, VR becomes, as Oppedisano (2020) defines it, a "paradigm for the knowledge of cinema culture" (115).

In this sense, the concept of "space for action," as developed by Colombo and Eugeni (1996), reveals to be useful. The two scholars distinguish the explicit space, which is seen on the screen when we act on it, from the implicit one, which is not seen when we look at the screen (or listen to the audio, we might add), but can become visible (or audible) if we perform certain actions in the explicit space. The space of perception-action is only the explicit space, where the implicit one is a space of possible action. The anchoring of the POV (and the POL) to a physical support, be it fixed or mobile, far from the being a limit, constitutes in this sense a necessary precaution to avoid getting lost in the space that we strive to consider 'real'. Moreover, as Gaudreault and Marion argue, every interactivity is intimately illusory (2013: 38-42), a contradiction that the very concept of virtual reality, as Maldonado points out (1994), express. The virtualization of reality, according to Maldonado, presupposes a series of questions about the materiality of reality and the immateriality of the virtual. The case of audio recordings, in this sense, may configure a philosophical and epistemological dilemma, as it entails a different integration between real and virtual experiences. Today, "technology is increasingly incorporated into the everyday fabric; it is transparent, wearable and ubiquitous, and therefore the poles of the man-machine union appear no longer dissociable but seem to form a new unit in continuous evolution" (Dalpozzo, Negri and Novaga 2018: 8). Dalpozzo (2018), in this sense, proposes a core question: should the current "early" virtual reality be considered only a form of playful entertainment? It is clear, to sum up, how the strong orientation towards the viewer, typical of the cinema of "attractions," re-emerges in VR cinema. Both devices, we may add, share the same purpose to educate audiences to a new image, characterized by an "attractive" aesthetics, which find a new ethical dimension in some contemporary productions. The same thing applies to audio, if we take up Chion's suggestion to overturn the relationship between audio and video within the cinema, where the concept of " 'natural harmony' between sounds and images" is always questioned by the postproduction process (2019: 98). The author reminds us that the 'audiovisual contract' (a notion which could be extended to any audiovisual medium) is "a kind of symbolic contract that the audio-viewer enters into, consenting to think of sound and image as forming a single entity" (Chion 2019: 249).

The relationship between sound and image is therefore not one of harmony, but of mutual contamination. The same 'contractual' concept of cinematic audio can be operated for audio in VR, for which perhaps a media-archaeological perspective can be equally relevant. The questions raised by Arcagni in "The Eye of the Machine" (namely in the chapter "The Virtual Eye") could be overturned, to this extent, in "The *Ear* of The Machine." Does the machine hear? What does it feel? What does it allow humans to feel? And above all, an enigma that would fascinate Chion: to what extent is its "ear" dependent, somehow, on its "eye"? Here too, the relationship between the two senses, and between the production and reproduction technologies connected to them, is as old as cinema itself.

"Is there the possibility of achieving, as far as sound is concerned, a result similar to that already achieved for light with photography?" This was the question raised, in 1857, by the worker-typographer Èdouard-Léon Scott de Martinville (Coeuroy 1929: 13, my translation), who went down in history for having invented the Phonautograph, the first device capable of recording sound. In the mid-nineteenth century, the chase game between sound and photo-cinematographic experiments, that will characterize the following decades, was already at work. This tension still reverberates today, with the new VR devices. Nevertheless, when the first recordings of the human voice began to circulate for international fairs, they shocked audiences just like moving images did, if it is true, as Coeuroy reports, that at the first presentation of Edison's phonograph, on 11 March 1878 at the Academie des Sciences in Paris, the representative Puskas was accused by the academic Bouillaud of ventriloquy (Coeuroy 1929: 26).

Initially, the phonograph was initially conceived as an alternative to the typewriter in corporate communi-

cations, so much so that it met with firm resistance from stenographers (Steffen 2005). Yet, already in 1878, in his volume dedicated to "radio microphones and phonographs," Count Theodor du Moncel, complaining about the limits of the machine, hoped for its use in the musical field: "if it were possible to reproduce at will a concert by the famous Adelina Patti, what a precious instrument the phonograph would become!" (du Moncel 1880: 246, my translation). The author's desire manifests not only a desire of a musical nature, but also that of reviving what has already been lived, of creating what we could call a sound phantasmagoria: "the wax figures representing great men will not only be able to render a faithful image of one's features, but also to make them speak, and the illusion will be complete" (du Moncel, 1880: 246-247, my translation). Edison's phonograph, therefore, like the cinema, satisfied and at the same time stimulated the desire not only to preserve and revive a simulacrum of reality, but also and above all to make it relive anytime and anywhere.

It is therefore not surprising that Edison himself, presenting his Kinetoscope in 1894, clearly declared that he wanted to "do for the eye what the phonograph did for the ear" (Edison 1888). An important point emerges here: at the origin of the two phenomena of sound and visual recording there is a substantial production convergence, aimed at filling the hunger for sounds and images that characterized the growth of mass society. In a famous image of the Black Maria studio, built by Edison and William K. L. Dickson in New Jersey to produce films for the Kinetoscope, in front of the set we find a compresence of a Phonograph and a Kinetoscope (Meeker 1894): this representation is involuntary emblematic of a triangulation between an old studio, in the background, and the new forms of technical reproducibility, in the foreground. The first is illuminated with natural light, the second are connected to each other through electricity; the same, on the other hand, that connects them to the surrounding space. In those studios, obviously, practices and professions will gradually diversify, but here we have a sort of primitive and incisive representation of a forced coexistence between two different media.

It is with this paraphernalia and with this scheme, we could say, that Edison was preparing to conquer Europe. *Dickson Experimental Sound Film* (1985), the very first experiment in this sense, proves to be particularly significant at the proxemic level. The violinist plays in the exact centre of the picture, and everything else is forced to adapt to his performance: the huge pavilion on the left, who tries to record it, the dancers, who dance it, and the off-screen operator, who tries to follow its rhythm. It was the first of a long series of experiments (Hébert 2007: 104), which in France culminated, in 1902, in Alice Guy's *chronomegaphone*, produced by Gaumont. As a rudimentary assembly of a gramophone and a cinema, the *chronomegaphone* was used in the so-called *phonoscènes*: beyond the technological amazement, these shows began to exercise a factual role in the diffusion of music (Keazor and Wubbena see them as the ancestors of the music videos) (2010). We will see, in the next chapter, how the experiments on VR audio seem to resume, to a certain extent, the same spirit of these pioneers, a comparison that reveals to be heuristically relevant. What is radically different is the technology, being the algorithmic architecture that lays beyond the device a cultural artifact like the device itself. Early film studies, software studies and sound studies, in the analysis of VR audio, are therefore forced, more than ever, to deal, and to constitute a new framework. This is the goal of the next two chapters.

4   Software Technologies for VR Audio Processing: Some Case Studies

From their very beginning, VR or 360˚ videos were expressly conceived for an immersive use on a visual level, trying above all to respond to the problems of stitching and parallax. For this reason, while many companies have developed the first cameras, and many video production studios have bought these devices and pushed them to their limits, few have taken audio into equal consideration. Today, according to the law of supply and demand, hardware and software companies are slowly adapting, revealing a need that is not only commercial but also cultural and perceptive (I have previously argued that film studies software studies, sound studies strongly converge on this point).

Audio in VR is variously defined with different expressions, sometimes confusingly – which is another typical problem of new media studies. We often talk indifferently of binaural, spatialized, spatial, reflected, spherical, 3D, VR, object-based and resonance audio, despite the attempts to distinguish the meaning of those terms (see, for instance, Fig. 1). More defined is the distinction between binaural, surround and stereo: recording audio for VR spatial applications audio needs binaural technologies, that is when different audio signals are sent to
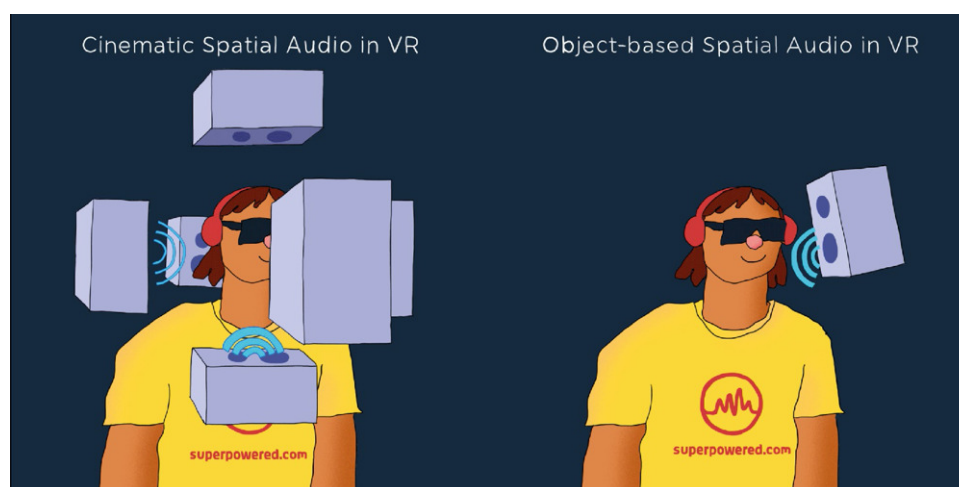
Figure 1. Audio VR scheme, divided into spatial and object-based ones (Source:
https://superpowered.com/3d-spatialized-audio-virtual-reality)

each ear to create a three- dimensional sound field (or "spatial sound"). Stereo is unable to create a natural and complete three-dimensional sound field because it is only able to position audio left or right, close or far from its original sources. Surround sound creates a more immersive aural experience than the stereophonic one; at the same time, it is still locked to a fixed, two-dimensional virtual environment. And yet, between binaural and stereo audio there is an intimate connection: just as in the leap from mono to stereo, space can change a musical mix, as it allows clarity and composition of sonic elements in a new way, transforming — at least potentially — its impact. Binaural is therefore perhaps the most exact term, although its uses are not limited to the field of virtual reality: binaural audio replicates the functioning of the human auditory system, considering the impulse responses of the brain and the so-called HRTF (Head-related transfer function), that is, a response that characterizes how an ear receives a sound from a point in space. Human anatomy, in other words, plays a fundamental role in the way in which sound waves are perceived, decoded and analysed by the human brain.

The standard, in this sense, is the Ambisonics sound, to which all hardware and software devices seem to adapt. Ambisonics is very different from traditional Stereo and Surround techniques; instead of sending the signal to predefined speakers, it treats the sound as a sphere that develops around a centre, which could be as much the position of the microphone during shooting as the position of the listener during playback. Thus, the signal is not dependent on any number or configuration of channels. The advantage of this format is the possibility of recording and decoding, at a later time, the audio material in another format, to rotate the whole sound seamlessly. Avoiding sudden and perceptible passages from one speaker to another, this standard can adapt to the most disparate needs.

Ambisonics is an encoding format: its information could be generated both by software and by microphones (as it is in the case of VR cinema, for instance). The microphones that allow to record in Ambisonics adopt several capsules, able to cover the sound sphere not only around the horizontal axis, but also to film above and below the listener. This is a focal point, which reveals how the additional dimension of height is able to recreate the feeling of being truly immersed in the experience. This method of creating three-dimensional sound fields later allows the POL (that is, where the player is) to be modified to any spot. This feature is what makes binaural audio the privileged solution for VR.

The head, with its mass and its "bulk," ensures that a sound that reaches the left ear first takes a few microseconds before reaching the right ear. The same concern, we may observe, has always been observed in concert audio systems, where the delay (or latency) is calculated in order not to give the effect of reverberation. The real innovation of spatial audio, indeed, is that the two sounds are enjoyed by the same person at the same time, trying to replicate the functioning of the human hearing system. When recording and creating audio for

spatial applications like VR, AR, XR creatives have to take all the rules and techniques of stereo recording and rethink them in terms of spatial application and 3D development. This rethinking is necessary both to give a (virtual) reality effect and to avoid that mismatch between the visually and auditory perceived environment that can be physically annoying, and cause effects of perceptual distortion and, ultimately, nausea. There have been numerous efforts, in both hardware and software development, to overcome these limits, in order to create a realistic sound effect (that is, essentially, an environment where "you can hear what you see"). We might try to analyse some of them, in order to understand, in the conclusions, how the overcoming of the limits constitutes, itself, a limit, in which a redefinition and a remediation of the spectatorial perception is at stake.

For what concerns the microphones, we might take into account the Zoom H3-VR virtual reality audio recorder, which is suitable for live streaming, video conferencing and direct recording. The microphone Soundfield SPS200 adaptively aligns the phases of the four microphone elements, which is useful for controlling timbre of different objects in recordings, or to control the balance between different objects. The Ambeo VR Mic is especially designed for 360° spatial audio recording, an Ambisonics microphone fitted with four matched KE 14 capsules in a tetrahedral arrangement, which allows to get fully spherical Ambisonics sound to match your VR video/spherical 360 content. All these microphones allow to capture complete spherical audio from a single point in space: basically, they record exactly what a listener would hear if he or she was in that position. To record the Ambisonics signal correctly, with regard to position and level, sounds easy, but some care should be taken, as certain mistakes cannot be corrected during post-production. To augment the number of microphones simultaneously at work seems to be helpful in reducing those mistakes: Core Sound OctoMic, for instance, have up to nine directional microphones, whereas the Zylia ZM-1 microphone contains some 19 mic capsules to the 4-8 above. In this last product, the idea is not necessarily VR: the company itself suggests to use this technology to automatically separate sound sources through one single device. In other words, if someone puts the "ball" in a studio recording, the software will separate out drums, keys, and vocalist. The parallel with VR cameras, such as Facebook Volumetric VR Camera, is evident, and follows, even visually, the line of the 'audio-visual chase' traced in the pre-cinema and early cinema years.

Figure 2. Comparison between ZYLIA ZM-1 microphone and Facebook Volumetric VR Camera (Sources: https://audioxpress.com/article/r-d-stories-the-genesis-of-the-zylia-zm-1-and-next-generation-audio-recording)

From the point of view of software production, while relying on the aforementioned Ambisonics technology, a trade war seems to be recurring between Facebook and Google (VR and spatial audio, in this sense, are just two battlefields in the wider context of a Big Tech competition). On the first side we have Facebook 360 Spatial Workstation: it is an end-to-end pipeline that allows sound designers to drop in audio sources, pan and sync to scene elements, and render to a single Ambisonic file that is played back on Facebook and Oculus video. Originally developed by Two Big Ears, Facebook 360 Spatial Workstation is now a free tool provided by the Audio 360 team at Facebook. It is a collection of plugins for Digital Audio Workstations (DAWs) that includes, for instance, a spatialised audio player and and a loudness meter. These plugins seem to be a pivotal element in

the technological chain where social values are encoded, at the interface between manufacturers and users, as they help authors to create spatial audio content, to encode it with platform-specific metadata (for Facebook, YouTube, etc.), and to play it back in a client application. On its side, Google recently launched Resonance audio – Audio Factory (Google), an open-source multi-platform, in the wake of the Chrome Experiments showcase, for all the creatives and engineers who deal with sounds in three dimensions. Spatial audio, here, could be edited and added to 360-degree videos, VR games and ARcontents. Resonance Audio, in other words, is a Software Development Kit (SDK) aimed at the developer community, based on the VR Audio technology of the Mountain View group and designed to allow the creation of experiences characterized by a high level of user immersion, even on mobile devices such as smartphones, where the computing power guaranteed by the processor is often limited. It is available on GitHub and supports various platforms for augmented and virtual reality, configuring itself as a sort of software development kit related to the reproduction of 3D audio. The realism achieved by Resonance Audio is entirely based on the relationships between sound waves, human ears and the type of environment in which the sounds are reproduced. Within a VR scenario, Resonance Audio is able to associate a sound with a specific source. To achieve this goal, the physical laws through which our hearing is able to localize a sound have been implemented in software: more specifically, in Resonance Audio there is a reproduction of the inter-aural delay for localisation on the horizontal plane of low frequencies (between 20 and 2,000 hertz) and the inter-aural difference of intensity for the localisation of high frequencies (between 2,000 and 20,000 Hz). All of these variances in how our right and left ears hear a sound allow us in real life to determine things like distance, height, and where a sound is originating from. Interaural time difference, in turn, has to be put in relation with the filtering effect of the HRTF, that is, the filtering of a sound source before it is perceived at the left and right ears.

Other software technologies operate on the same level. The Ambisonics Toolkit, for instance, is a free spatial audio plugin, which brings together a number of classic and innovative tools for the artist working with Ambisonics surround sound (it can be used, for example, with the GoPro VR Player). Envelop for Live (E4L) is a free set of tools provided for musicians who aim to work with spatial audio production and live performances. IFFMpeg is a tool that allows to export different video types/codecs for different platforms attaching the spatial audio to the video (the act of mixing/merging of audio and video is called 'muxing'). Dear VR allows the users to mix immersive music, games, VR, AR and 360° video productions in a simulated DAW, supporting binaural, Ambisonics, and multi-channel loudspeaker output formats at once. Drops is built to provide a way for those with little or no musical background or training to create rich and expressive polyphonic rhythms. Tribe XR is a tool useful for learning to be a DJ in VR. G'Audio leverages object-based mixing for detailed placement of sound in a 3D environment, with a plug-in, G'Audio Craft, especially designed for gaming sonic experiences. Steam Audio is an audio package that can combine multiple occlusions, reflections, reverbs and effects (which, in turn, may work with other external software, like V-B Audio Voicemeeter Banana for Windows). Other plugins are available for audio or video production applications such as Avid, Première, Reaper, LiveSwitch. Other tools to create spatial music are are The Music Room (an expressive MIDI controller), Lyra VR (which allows to interact with 3D music sequences) and Electronauts - VR Music (an environment where one build, drop, remix and jam with friends and artists). Particularly relevant, for a comparison with the video technique, is GoPro Fusion Spatial Audio (GoPro rigs has been one of the first, most cheap and most used VR cameras, before the development of cameras such as Ricoh Theta, Kodak SP360, Giroptic 360cam, IC Real Tech Allie).

All of the apps mentioned amply demonstrate the potential of VR sound and music making, introducing creative tools that overcome old limits and generate new ones, where the two phenomena seem to be intrinsically connected. And yet, the feeling is that many of these ideas need to be honed and polished by creatives and professionals. Furthermore, many of the so called "innovations" seem to be intimately old, and connected to a 19th Century technological scenario. For instance, Ossic (formerly SonicVR) is developing the Ossic X, a highly sophisticated headphone in which every component — including head tracking — is designed for immersive 3D-audio playback. This algorithm, like others developed with Ambisonic audio codecs, synchronize the spatial audio channel with the head position, according to three combined parameters: dialogue with the viewer, integration of the viewer within the audiovisual fiction and three-dimensional simulation. Has cinema ever done anything else? Or, if we consider again Google's audio resonance patterns, we might see (Fig. 3) how the "bounce" patterns of sound frequencies are similar to those of light frequencies. Majkut's separation between

centre of orientation and *corner of orientation* seems to have, here, an emblematic representation, directly linking epistemological distinctions to software and hardware features. This epistemological clash entails a certain degree of complexity in the research on audiovisual immersion: beyond the apparent simplicity of the scheme proposed, indeed, Resonance Audio hides an obsessive attention to detail (the full project even consider and calculate the bounce of sound waves in the helix of the ear). It is precisely in this obsessive research for a fully immersive experience that VR technologies, both in the fields of video and audio production, turns out to be illusory and, as such, "non-immersive," as they constantly recall the ability of the device to "attract" through its technological features. From this consideration, I will try to sketch some similarities with early cinema practices in the next chapter, testing the framework proposed in the previous ones.

Figure 3. Resonance Audio SDK for Unity (Source:
https://forum.unity.com/threads/resonance-audio-sdk-for-unity-deliver-high-fidelity-spatial-audio-at-scale.503347/)

5   Conclusions

As we have seen, to compare moving images and spatial audio, just like *sound studies* attempted to do in the past decades, reveals to be fruitful, even the field of VR. Compared to traditional cinematic language, indeed, the difficulties that spatial images face are, in short, the absence of a hierarchy of plans and the impossibility of modulating the light environment without, for this, being forced to exhibit the apparatus. If this second aspect doesn't seem to be relevant from an audio perspective (where intradiegetic/non-diegetic distinction remains unchanged), the hierarchy of listening plans, on the contrary, seems to be a field in which engineers and creatives are jointly trying to work. Even only to decide whether to interact with a spatialized audio or not turns out to be a difficult question: taking into account the contribution of sound studies, as Paul Carter (2004) notes, "auditory space is durational, but it lacks music's (and writing's) commitment to linear development. Without a sense of ending, it is not located between silences" (59). What is the role, then, of spatial audio? Will it develop in a combination of directional music and non-directional audio? To what extent might the choice of a listening point be the responsibility of the spectator (or audience)? How much is that choice driven by the director, how much by software technologies, how much by the hardware devices?

The scenario seems indefinite, even more if, as we have tried to sketch, we consider the software as something which encloses a variety of products: protocols, plugins, proprietary coding systems (all these topics as been threated, as we have seen, by software studies). These products seem to be relevant for the values that they embed, the practices that they foster or that they are part of, in multiple and indefinite ways. It is precisely this indefiniteness that prompts us, both in the audio and in the video domain, to consider VR as a reconceptualization of our visual and auditory grammars of everyday life. We may see, as well, as those grammars are determined by a series of complementary and irreducible factors. In this complexity, we could say, the audiovisual aesthetics of the future is being and will be determined, at least potentially, by new media artists, which are bringing this expressive form to its structural limits. Spatial audio, therefore, seems to reveal, in the light of the case studies analysed, the importance of the devices and techniques, understood in a cultural and social sense, as determining factors with their own agency, and their own ideology: in this sense, as well, VR

may be intended in continuity with the history of cinema, and represents (or can be interpreted as) one of its multiple manifestations.

References

Andreoli, Marco (2018). "La realtà virtuale al servizio della didattica." *Studi sulla Formazione* 21: 33-56. https://doi.org/10.13128/Studi_Formaz-23090.

Amoruso, Giuseppe (2019). "Digital Technology for Knowledge, Design and Experiential Education for Culture." In *Digitalization and Cultural Heritage in Italy. Innovative and Cutting-Edge Practices*, edited by Fernando Salvetti and Antonio Scuderi, 12-22. Milano: Franco Angeli.

Arcagni, Simone (2020). *Immersi nel futuro. La realtà virtuale, nuova frontiera del cinema e della TV*. Palermo: Palermo University Press.

Arcagni, Simone (2018). *L'occhio della macchina*. Torino: Einaudi.

Basso, Alessandro (2020). *Ambienti virtuali per nuove forme di comunicazione*. Canterano: Aracne.

Battini, Carlo (2017). *Realtà virtuale, aumentata e immersiva per la rappresentazione del costruito*. Firenze: Altralinea.

Brinkman, Willem-Paul, Allart R. D. Hoekstra and René van Egmond (2015). "The Effect of 3D Audio and Other Audio Techniques on Virtual Reality Experience." *Studies in Health Technology and Informatics* 219: 44–48. https://doi.org/10.3233/978-1-61499-595-1-44.

Çamcı, Anil and Rob Hamilton (2020). "Audio-first VR: New Perspectives on Musical Experiences in Virtual Environments." *Journal of New Music Research* 49(1): 1–7. https://doi.org/10.1080/09298215.2019.1707234.

Carlo, Simone and Fausto Colombo (2007). "La digitalizzazione. Questioni strutturali." In *La digitalizzazione dei media*, edited by Fausto Colombo, 15-38. Roma: Carocci.

Carter, Paul (2004). "Ambiguous Traces, Mishearing, and Auditory Space." In *Hearing Cultures. Essays on Sound, Listening and Modernity*, edited by Veit Erlmann, 43–63. New York: Berg.

Casetti, Francesco (2019). "Primal Screens." In *Screen Genealogies. From Optical Device to Environmental Medium*, edited by Craig Buckley, Rüdiger Campe and Francesco Casetti, 27-50. Amsterdam: Amsterdam University Press.

Chion, Michel (2019). *Audio-Vision: Sound on Screen*. New York: Columbia University. 1st ed. (1990). *L'audiovision*. Paris: Nathan.

Chun, Wendy Hui Kyong (2016). *Updating to Remain the Same. Habitual New Media*. Cambridge, MA - London: The MIT Press.

Colombo, Fausto and Ruggero Eugeni. *Il testo visibile. Teoria, storia e modelli di analisi*. Milano: Carocci.

Coeuroy, André (1929). *Le phonographe*. Paris: Kra.

Cook, Matt, Zack Lischer-Katz, Nathan Hall, Juliet Hardesty, Jennifer Johnson, Robert McDonald and Tara Carlisle (2019). "Challenges and Strategies for Educational Virtual Reality. Results of an Expert-led Forum on 3D/VR Technologies Across Academic Institutions." *Information Technology and Libraries* 38(4): 25-48. https://doi.org/10.6017/ital.v38i4.11075.

Dalpozzo, Cristiano (2018). "Cinema e realtà virtuale, ovvero 'the early virtual (post)cinema of attractions'." In *Dalpozzo, Negri, Novaga 2018 (eds.)*, 87–106.

Dalpozzo, Cristiano, Federica Negri and Arianna Novaga (eds.) (2018). *La realtà virtuale. Dispositivi, estetiche, immagini*. Milano - Udine: Mimesis.

du Moncel, Théodose Achille Louis (1880). *Le microphone, Le radiophone et le phonographe*. Paris: Hachette.

Edison, Thomas Alva (1888). "Patent Caveat 110." West Orange, NJ: Edison National Historical Site Archives, 8th October.

Fletcher, Connor, Vedad Hulusic and Panos Amelidis (2019). "Virtual Reality Ear Training System. A study on Spatialised Audio in Interval Recognition." In *11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. https://10.1109/VS-Games.2019.8864592.

Galloway, Alexander R. (2004). *Protocol. How Control Exists after Decentralization*. Cambridge, MA: The MIT Press.

Galloway, Alexander R. (2012). *The Interface Effect*. Cambridge, MA: Polity.

Gamper, Hannes and Tapio Lokki (2010). "Audio Augmented Reality in Telecommunication through Virtual Auditory Display." In *The 16th International Conference on Auditory Display (ICAD-2010)*, 63–71. http://hdl.handle.net/1853/49841.

Gaudreault, André and Philippe Marion (2013). *La fin du cinéma? Un média en crise à l'ère du numérique*. Paris: Armand Colin.

Gibson, James J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Hébert, François (2007). *Dans le noir du poème: les aléas de la transcendence*. Montréal: Fides.

Jung, Timothy et al. (2020). *Augmented Reality and Virtual Reality. Changing Realities in a Dynamic World*. New York: Springer International Publishing.

Keazor, Heazor and Thorsten Wübbena (2010), *Rewind, Play, Fast Forward. The Past, Present and Future of the Music Video*. Bielefeld: Transcript Verlag.

Kern, Angelica C. and Wolfgang Ellermeier (2020). "Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence." *Front. Robot. AI*, 21st February. https://doi.org/10.3389/frobt.2020.00020

Kim, Hansung, Luca Remaggi, Philip J.B. Jackson and Adrian Hilton (2019). "Immersive Spatial Audio Reproduction for VR/AR Using Room Acoustic Modelling from 360° Images." In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 120–126. https://doi.org/10.1109/VR.2019.8798247.

Lanier, Jaron (2017). *Dawn of the New Everything. Encounters with Reality and Virtual Reality*. New York: Henry Holt and Company.

Lefebvre, Henri (1974). *La production de l'espace*. Paris: Anthropos.

Luigini, Alessandro and Chiara Panciroli (2018). *Ambienti digitali per l'educazione all'arte e al patrimonio*. Milano: Franco Angeli.

Majkut, Paul (2014). *Smallest Mimes. Defaced Representation and Media Epistemology*. Bucarest: Zeta Books.

Maldonado, Tomás (1994). *Lo Real y Lo Virtual*. Barcelona: Gedisa.

Manovich, Lev (2013). *Software Takes Command*. London - New York: Bloomsbury.

Manovich, Lev (2001). *The Language of New Media*. London - Cambridge, MA: The MIT Press.

March, Ramiro Javier (2001). "Information, image, réalité virtuelle et réalité." *Archeologia e Calcolatori* 12: 275–305.

McQuire, Scott (2008). *The Media City. Media, Architecture and Urban Space*. Los Angeles: Sage.

Meeker, Edward (1894). "Century Magazine Interior of the Kinetographic Theater." Orange, NJ: Edison's Laboratory, 1st June.

Nau, Jeffrey R. (2019). *Virtual Realities in Archaeology. Employing the Oculus Rift for Artifact Visualization and Education*. Master's Theses. Kalamazoo, MI: Western Michigan University.

Oppedisano, Federico Orfeo (2020). "Le strategie immersive del cinema tra attrazione e narrazione." In *Rossi 2000*, 103–118.

Parisi, F. (2020). "La sintonia sensomotoria nella realtà virtuale." *Reti, saperi, linguaggi. Italian Journal of Cognitive Sciences* 1: 85-102. https://doi.org/10.12832/98420.

Rivette, J. (1961). "De l'abjection." *Cahiers du cinéma* 120: 54–55.

Rossi, Daniele (2020). *Realtà virtuale. Disegno e design*. Canterano: Aracne.

Rovithis, Emmanouel et al. (2019). "Bridging Audio and Augmented Reality towards a new Generation of Serious Audio-only Games." *Electronic Journal of e-Learning* 17: 144-156. https://doi.org/10.34190/JEL.17.2.07.

Rovithis, Emmanouel et al. (2019). "Audio Legends: Investigating Sonic Interaction in an Augmented Reality Audio Game." *Multimodal Technologies Interact* 3(73). https://doi.org/10.3390/mti3040073.

Schafer, R. Murray (1977). *The Tuning of the World*. New York: Random House Inc.

Steffen, David J. (2005). *From Edison to Marconi. The First Thirty Years of Recorded Music*, London: McFarland & Co.

Steffen Jacob H, James E. Gaskin, Thomas O. Meservy, Jeffrey L. Jenkins and Iopa Wolman. (2019). "Framework of Affordances for Virtual Reality and Augmented Reality." *Journal of Management Information Systems* 36(3): 683-729. https://10.1080/07421222.2019.1628877.

Steincke, Frank (2016). *Being Really Virtual. Immersive Natives and the Future of Virtual Reality*. New York: Springer.

Sterne, Jonathan (ed.) (2012). *The Sound Studies Reader*. London - New York: Routledge.

van Helvoort, Daniël, Emil Stobbe, Richard Benning, Henry Otgaar and Vincent van de Ven (2020). "Physical Exploration of a Virtual Reality Environment. Effects on Spatiotemporal Associative Recognition of Episodic Memory." *Memory & Cognition* 48(4): 691–703. https://doi.org/10.3758/s13421-020-01024-6.

Raffaele Pavoni – University of Florence (Italy)

✉ raf.pa53@gmail.com

Raffaele Pavoni earned a PhD in Storia delle Arti e dello Spettacolo, at the University of Florence. The topic of his doctoral research is Italian music videos contemporary production and consumption. In his work, based on ethnographic analysis and market data, he tried to draw the actual institutional framework of music videos, and to understand how their transition to the web has changed the relation between audience, directors, music labels and software houses. Actually, he is studying the representations and self-representations of migrants in the old and new media, with a research project that intertwines film studies and visual anthropology.