Paper published as:

Author/s: Candusso, D.

Title: Designing spatial sound: adapting contemporary screen sound design practices for virtual reality

Conference Title: SMPTE17: Embracing Connective Media

Dates held: 18/07/17 – 20/07/17 **Place**: Sydney, Australia

Pages: 1-10

Abstract:

The introduction of 360º film and Virtual Reality (VR) has meant practitioners have had to adapt to a multitude of new platforms and specifications, with no existing pathways or working methodologies. VR audio is continually being redefined, with contemporary film sound practices having to adapt to new forms of spatialisation and headphone delivery. The variations in platforms also necessitates different delivery requirements, with the audio often a secondary consideration. This paper will examine the current state of sound for virtual reality and discuss sound design for VR from a creative practice perspective.

DOI/Link: http://dx.doi.org/**10.5594/M001749**

Designing Spatial Sound: Adapting Contemporary Screen Sound Design Practices for Virtual Reality

Dr Damian Candusso M.P.S.E, Charles Sturt University

Charles Sturt University, Boorooma St, Wagga Wagga, dcandusso@csu.edu.au.

Abstract. *The introduction of 360º film and Virtual Reality has meant practitioners have had to adapt to a multitude of new platforms and specifications, with no existing pathways or working methodologies. Audio for Virtual Reality is continually being redefined, with contemporary film sound practices having to adapt to new forms of audio spatialization and headphone delivery. The variations in virtual reality platforms also necessitates different delivery requirements, with the audio often a secondary consideration. Sound designers and editors are either delivering a non-spatial soundtrack, or they are having to adapt to various spatial plugins and/or game engine platforms in the creation of a cohesive spatial soundtrack. This paper will examine and discuss the current state of sound for Virtual Reality from a creative practice perspective.*

Keywords. *Sound Design, Virtual Reality, Virtual Reality Sound, 360º, Film Sound, VR, VR Sound, Immersive Sound, Immersive Sound, Immersive Audio, Game Sound, Game Audio, Binaural, Ambisonics, HMD, Interactive Film.*

Introduction

In the tradition of live-action filmmaking, it has usually been that the image is captured before the sound. This sequence of production stems from the beginnings of cinema when it was only the visual that could be captured due to synchronization limitations. Although the coming of sound brought with it a new dimension to filmmaking, the emphasis has always favored the visual. Doane notes that going to see a film, affirms the denial of heterogeneity between image and sound. 'Sound is something that is added to the image, but nevertheless subordinate to it - it acts, paradoxically, as a "silent" support.' (Doane,1985, p. 54)

Film technologies are improving and elements of contemporary filmmaking are challenging the methodology of the image preceding the sound. This is most notable with animated films that often begin their production process with the dialogue script recorded first and the animation following. With a film only taking on life through collaboration, this approach to sound and vision is becoming more important as films become increasingly fabricated and/or digitally enhanced. Sound for animation has the power of transforming stilted imagery, thus creating the illusion of fluidity and reality. Although it is undoubtedly true that sound contributes to live-action film, it is the contribution that sound makes to animation, computer generated special effects and now Virtual Reality (VR), that sound yields even greater results. Sonnenschein recognizes that sound can provide the impression that it is contained within the image when it is derived from the interaction between the natural world and our psychological responses. (Sonnenschein, 2002, p. 168) The introduction of VR allows sound to go beyond being simply contained within the image. The field of view through a VR headset is restricted to 110º, however if the soundtrack is created to its full potential, it has the ability to be a full spherical 360º around the viewer.

360º Film and VR

It is important to differentiate between 360º video experiences and Virtual Reality. 360º videos essentially wrap a screen around you, and as the viewer, you are able to turn your head all around to experience the film/imagery. VR on the other hand, allows you to interact, move and walk through the environment changing your perspective to both the visual elements and the sound sources. Although 360º video also allows you to change your perspective, it is restricted to head rotation only.

In this paper, I will refer to both VR and 360º video as they often present similar challenges in relation to soundtrack creation. VR is still in its infancy and it is being driven by a common desire to create an immersive experience with technical developments and workflow practices continually evolving. No one is really an expert, with this medium largely unchartered as it is being redefined daily. This is exciting times for those working in sound, as according to Foder, unlike any other creative medium, people are pushing the importance of sound from top to bottom. (Foder, 2016) The industry is beginning to adapt to immersive sound, however, it requires educating all involved so that it can be a priority consideration in VR production.

Virtual Reality represents immersion and immersion represents the user being present in a virtual world. 'Presence is a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience.'(International Society for Presence Research, 2009) It is the combined level of an homogenous audio-visual experience that is providing the greatest hurdles to a truly immersive experience. As discussed earlier, Doane's notion of the dislocation

between the sound and image of a film is also true for VR. Mona Lalwani states ‘that the premise of VR is to create an alternate reality, but without the right audio cues to match the visuals, the brain doesn't buy into the illusion. For the trickery to succeed, the immersive graphics need equally immersive 3D audio that replicates the natural listening experience.’ (Lalwani, 2016) Alex Riviere from Audiokinetic, the creators of audio middleware Wwise, adds further to this notion by stating that sound is ‘perhaps the most powerful medium we can develop in order to “hack” the user’s brain and create a complete 360-degree experience that offers spatial immersion.’ (Riviere, 2016)

Despite hearing being one of the first senses we develop at about 4 months of age, whilst still in the gestation period in our mother’s womb, it remains one of our least understood senses. However, as indicated by Balazs, our ears can often provide far greater detail than our eyes. ‘The number of sounds and noises a human ear can distinguish runs into many thousands¬ far more than the shades of color and degrees of light we can distinguish.’ (Balazs, 2015, p. 212) Victor Phoenix from Technicolor sees the invisible contribution of sound as a sign of success. ‘VR is an exciting new way to tell stories. It’s looking at the history of cinema and games and then seeing how we can tell new stories in this exciting new way. But just like film or any other form of entertainment, when you don’t notice it, when you don’t think about [it], when you’re just deep in the virtual environment, and when the sound is invisible, that’s when it’s at its best—when it’s simply part of the cohesive experience.’ (Baker, 2017)

The evolution of screen based sound technologies has demonstrated that with each new technological advancement has come an increase in the number of speakers used. Generally speaking, we have evolved from mono to stereo, 4-channel surround to 5.1 and 7.1 surround to the new immersive sound formats including the 128-track Dolby Atmos and Barco Auro that include height channels. However, alternative formats including ambisonics and binaural sound through headphones also have the potential to get us closer to a 3-D soundtrack.

For Virtual Reality to be effective, we need to be placed inside an alternative reality. This requires a Head Mounted Display (HMD) and headphones to remove and isolate us from the real world. One of the greatest developments with HMD has been the ability to allow head tracking. Not only can the participant turn their head to look wherever they want, the soundtrack can also follow the HMD orientation allowing sounds to remain bound to the imagery.

360º video also allows a HMD to navigate a spherical world with a matching soundtrack. However, as the soundtrack is often a single ambisonic file created through conventional film post production workflows, the soundtrack has no real-time rendering. Instead, the panning of the entire soundtrack changes, but there are no dynamic variables including volume. This also highlights the importance of sound source and sound space considerations. Designing sound for immersive media requires an understanding of sound in terms of space and acoustics. The entire film and VR industry need to be educated on this, as one of the greatest challenges facing VR at present is a lack of understanding about spatial sound. Studios are producing and releasing 360º experiences, however, they are often using traditional post production techniques, continuing to think in terms of speakers and speaker placement. These studios are used to working in formats such as 5.1 and 7.1, however, by continuing to apply the exact same channel based work practices, they are limiting the possibilities and benefits of spatial audio, and this interrupts the viewers immersive experience. Traditional approaches do not work as the listener is now in a sphere. Speaker orientations do not work as the sound orientation/ spatialization is derived completely from head tracking and a sound can be anywhere, not just from the left, centre or right channels etc. Creating a static stereo mix or a 5.1 mix is selling the experience short, with all VR platforms supporting a full 3-D head-tracking soundtrack.

There is no doubt that increasing audio realism will enhance the immersive experience and engagement levels. However, what we are seeing are new challenges as a result of this shift from traditional linear sound to immersive sound. Currently there is limited documentation on sound design within the VR context. As such, there are no rules or standardizations associated with this new practice. As Kuzminski notes, 'virtual reality is a new frontier for audio and sound design. The industry is buzzing with potential and new technology, though many elements have yet to be figured out or agreed upon.' (Kuzminski, 2016)

In the early 2000s games and film were seen as competing mediums, and there was little intellectual property being shared between these two industries. Fast forward to today, and it is becoming increasingly obvious that both mediums are beginning to merge with the line between linear film, interactivity and gaming beginning to fade. By adding a game engine to the workflow, object based sound and real-time processing can be used to create a far more realistic and creative soundtrack.

Game sound engines work by placing sound *objects* within a 360º world, whilst allowing these sounds to interact and take on acoustic properties derived from their location and environment. This includes reverberation and occlusion characteristics. The introduction of room scale VR also allows the user to walk around a space including the ability to walk up close to objects. In our real world, we know that by moving around a space our acoustics change relative to the space, and this can also be created within the game engine in real-time.

As Sound Designers, we now need to re-think the use of the sound space. Through binaural sound, it is possible to simulate three-dimensional positioning of sound sources and to have sounds envelop the user. VR places the user *inside* the virtual world and it is here where the soundtrack has the power to reflect reality, and the acoustics of our real world. 'Since the visuals aren't perfect imitations of reality, the audio is vital for maintaining immersion and supporting the game play. Details in the audio actually help the brain process the visuals faster.' (Walden, 2016) The acoustic environment has such an impact on this replication that controlling these variables creates challenges for VR audio and sound designers. This is further complicated when there is a need to create a spatial connection between the soundtrack and the 360º and/or 3D VR imagery; specifically, creating a true 3-D soundscape that surrounds the participant and supports the image.

The process

Sound needs to be an early consideration in any immersive project. One of the first questions that needs to be asked is, what platform is the project delivering on (YouTube, Facebook, Oculus, GearVR, Vive etc.)? With different technologies and delivery platforms in competition with each other, every platform requires a separate approach. 3-D sound is essential to providing a fully immersive experience and it is used to control the narrative. For example, your field of view is limited to the viewing range of each specific device at around 110º, however a well-crafted soundtrack will provide a full 360º immersive sound field with positioned objects. It is here the power of the soundtrack comes to the forefront as the sound is now able to control the VR viewer – the sound directs the viewer to follow the narrative and to tune in to specific cues. Audio should be created with sound objects that are located corresponding to the player's location and head movements.

Audio working tools continue to be developed in order to enhance the VR experiences and for improving the required new mixing techniques. One of the first areas to see developments has been in 360º audio positioning. Software including Facebook's 360 spatial workstation and Dolby Atmos VR tools provide linear timeline based object panning for sounds in 360º video, before being rendered to an ambisonic file. Due to the entire 360º video soundtrack being a

rendered file, a limitation to the format is that the only variable is the ability to pan the entire soundtrack in relation to head rotation.

In many ways, these technologies are playing catchup to the foundations of game engines where their *entire* working ecosystem is built upon objects. The object paradigm translates well with binaural audio, and suits the nature of VR as it requires headphone playback. This process is typically implemented by taking each individual audio object, and passing it through position-dependent filters called Head-Related Transfer Functions (HRTF). These filters model the interaction of sounds with the head, and can succeed in giving the impression that sounds of the game are actually coming from inside or outside the player's head.

Despite much attention directed at improving the positioning of sound within the 3-D space, another major implication currently challenging industry is the way these sounds interact and behave in relation to this virtual environment. That is, the sounds also need to change based upon their interaction with other objects and the acoustics of the space. All day, our ears and brain interpret complex audio interactions that change based on the shape, the material the reverberations, reflections and diffractions of the environment around us. In addition to providing feedback on the source of sounds, it also provides an aural environmental map.

Technologies and workflows

In summary, there are two distinct methods for working in VR and 360º. True interactive VR needs to be created within a game engine, whereas 360º can be created in either a game engine or within a DAW. The table below (Table 1) highlights some of the available technologies.

Table 1.
Comparison of available technologies for VR and 360º audio production

VR (interactive) + 360º	**360º (Linear)**
Game Engines	**DAWs**
Unity	Protools
UE	Reaper
	Nuendo
Middleware (although not needed, it does often make working easier)	
Wwise	
FMOD	
Spatialization Plugins	**Spatialization Plugins**
Oculus	Dolby Atmos
Steam	FB360 Spatial Workstation
	Several other 3rd party plugin option

Linear sound

Soundtrack creation for linear media follows a path where the sound elements are recorded, edited, mixed and delivered as fixed channel-based outputs. The diagram below (Figure 1) shows a typical workflow for traditional linear media.

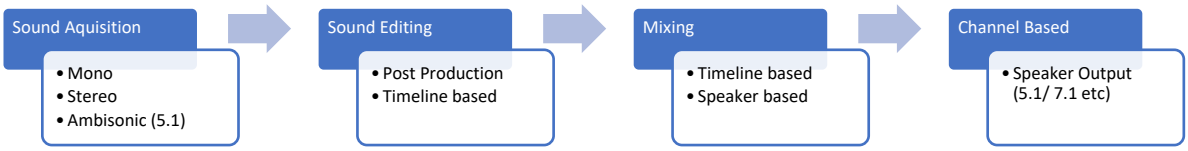

Figure 1. Traditional Linear Media Workflow

Immersive cinema formats

The introduction of immersive cinema formats has allowed sound objects to be included in addition to the traditional channel-based formats (Figure 2). A very important distinction of channel-based audio is it relies on a fixed number of channels from the point of production to playback, and audio mixes are designed for a strict speaker configuration or have to be down-mixed to accommodate simpler speaker set-ups. Object-based audio allows sounds to have a 'position' and it is this position that is rendered at time of playback dependent on the individual speaker configuration of each cinema. The objects play along the timeline as encoded streams, and the meta-data within the stream provides 3D positional data that is then decoded to the physical speakers.

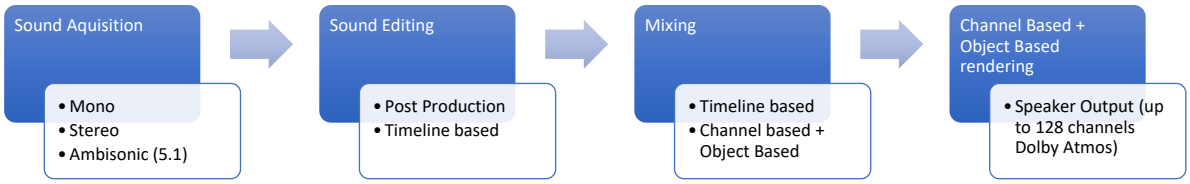

Figure 2. Film Object Based Rendering Sound Workflow

360º Video

The workflow for 360º video allows objects to be positioned within a 3-D space, however this is then rendered to an ambisonic file that allows the entire soundtrack to pan and orientate with head tracking (Figure 3).

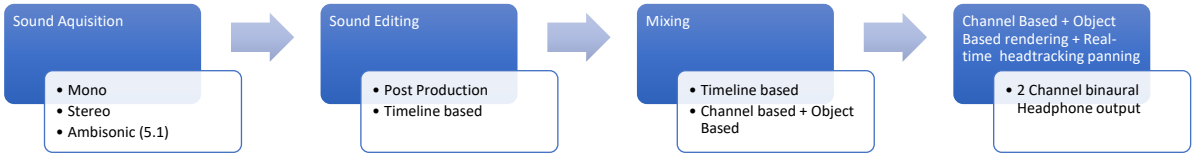

Figure 3. 360º Video Sound Workflow

Virtual Reality Sound Workflow

Working in Virtual Reality requires a more complex workflow. This includes creating the soundtrack natively within a game engine or by using 3rd party middleware linked to a game engine. In addition to objects being positioned anywhere in 360º, game engines offer powerful real-time processing including mixing, panning, volume attenuation and acoustic spatialization that can be implemented to improve acoustic emulation (Figure 4).

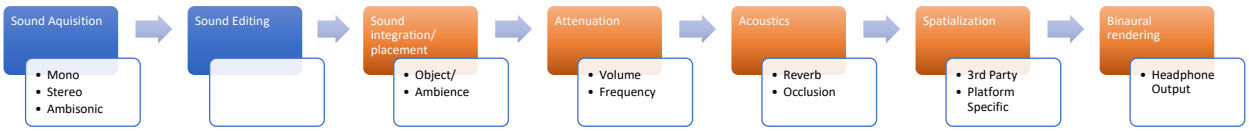

Figure 4. Non-Linear Virtual Reality Sound Workflow. Note that there is real-time processing from 'Attenuation' through to Binaural rendering

What implications are there for the sound designer?

The transition to working in VR has exposed many challenges for sound designers. Of significance is the jump that traditional linear sound editors and mixers will have to make to a platform that relies on sequenced objects. However, the rewards are worth the steep learning curve as the VR community understands that better audio realism increases the immersion into the VR narrative.

There is limited documentation and currently no standardizations regarding how best to work in VR. Questions are constantly being asked about how best to record, edit, mix and present work. The mixing tools are still very much game orientated and the lack of a user-friendly GUI can discourage film sound designers. However, they need to embrace the medium and these new tools that allow greater flexibility for spatial sound design in a 3-D environment. Mixing practices need to adapt to the new technologies and formats as there is no longer a focus on speakers or channels. The dialogue should now pan as we don't have a fixed screen and a center channel. Sound designers need to re-think the use of the sound space. There is a need to create a spatial connection between the soundtrack and the imagery by creating a true 3-D soundscape that surrounds the participant and not only supports the image, but also leads the image. Head tracking is an additional layer of mixing, so thought needs to be applied to determine what needs head-tracking and what needs to remain fixed in mono or stereo. For example, sound objects versus a voiceover inside the listener's head. Extra time is needed for these details, and this needs to be budgeted for.

Sound designers need to get involved early in the production process. It may be necessary to be on-set and on-location to record in specific formats including binaural and ambisonics, or a hybridized approach to sound acquisition. For example, this could include ambiences and atmospheres recorded in ambisonics, with additional character dialogue individually miked. Sound designers are no longer just working to linear imagery, instead it is now a hybridization between linear and interactive. The use of game engines introduces new challenges for traditional linear sound designers. Real-time rendering relies on processing power so sound designers need to now have some understanding of the limitations of the technologies being used. For example, editing and mixing practices may need to adapt as necessary to accommodate memory and processing limitations.

In 2016, the VR industry generated an estimated total revenue of $1.8 billion, and it is expected to continue to boom. (Ansari, 2017) Research firm the International Data Corporation (IDC) forecasts worldwide revenues for the augmented reality and virtual reality (AR/VR) market to reach $13.9 billion in 2017.(International Data Corporation, 2017) This is great news for the sound industry, however working practices need to adapt. Marcie Jastrow, Technicolor's Senior VP of Immersive Media and Head of the Technicolor Experience Center, says 'the biggest challenge is going to be learning who the new storytellers will be. What is often written out for creative is not something that can translate into immersive experiences.' (Jastrow, 2016) The relationship with technology is going to strengthen as the sound designer is no longer an editor

and/or mixer; instead, they now need to understand how game engines, game audio middleware, interactivity, scripting and coding work. This leads to a knowledge of being able to integrate various sound formats, not just to create a sonically rich soundtrack, but also how to combine these sounds and formats to maintain spatial integrity. There is a difference between sound for live-action VR and animated CG VR and there is often a hybridity of both. In addition to capturing spatial sound on set, what sounds can be fabricated through tools? Each platform provides many unique challenges.

As an emerging field, VR is constantly changing due to the reliance on technology. Sound designers need to be able to adapt to change, and embrace this technological evolution of software and working methods, with the underlying requirement to continue to tell a story regardless of medium. Sound Designers need to keep abreast of emerging technologies, and how best to utilize these. Certainly, in the short term, they need to be able to jump across multiple applications for sound acquisition, editing, mixing, spatialization and final output. In addition, the sound professional working in VR needs to increase their understanding of coding and collaboration with IT professionals. Particularly if working within a game engine, the sound designer is often required to implement their sounds to allow interactivity. It is far more efficient for the sound designer to implement these without having to rely on coders.

Essential to good VR sound work practices is the ability to incorporate the final delivery platform into the workflow. For example, if delivering for the HTC Vive, it is essential that the sound designer be able to audition work within the Vive, and then be able to seamlessly make changes and alterations. Working in VR requires the sound designer to move back-and-forth as they constantly need to work between the game engine, audio editing software and the final delivery platform. This requires a constant change to the work environment, going from working on a flat screen and possibly speakers to wearing a headset and headphones, and then swapping back again. Each VR experience is different and different VR producers are using a different set of tools.

With so much dependence and change with VR technology, it is essential that the sound design process be well planned in advance. As noted by technicolor's Marcie Jastrow, 'when you are building immersive experiences, there is often a nine-month lead time. A lot of new innovations in technology can emerge in that time frame. So, you need to have a crystal ball to see what technology is going to be released and figure out how you can push it forward.' (Jastrow, 2016)

Conclusion

Creating sound for 360º video and Virtual Reality requires adapting contemporary film sound practices in addition to the incorporation of game sound techniques. Sound needs to be an early consideration for VR projects so that the sound designer has the ability to record on-set, and be able to determine the workflow that best suits the final release platform and all available technologies. These new mediums are allowing for the use of non-traditional recording techniques including ambisonic and binaural, however these only provide a solid bed and so much more can be added. No longer is the sound designer working to a linear narrative and timeline with fixed speaker channels; instead they have the ability to create a fully immersive and interactive 360º acoustic experience.

360º immersive sound is not just limited to placing sounds around a 360º environment. Through the incorporation of a game engine, real-time rendering offers many advancements to the creation of a realistic player based virtual world. More than ever, accurate acoustic modeling is an important factor including the volume balance, spatial placement of sounds, and reverb and

frequency characteristics as a result of real-time occlusion. Game engines also offer the ability to start and stop sounds based on character location and timeline cues.

Due to VR platforms having different specifications, limitations and delivery formats, one of the challenges of creating sound for immersive media is having the access to the final release platform. This is the only way that a sound designer can create the best soundtrack possible through editing, designing, mixing, and spatializing effectively.

References

Ansari, I. (2017) 'Super Data Revealed Number of Sales and Revenue Charts For PSVR, Rift, GEAR, VIVE', SegmentNext, 3 February. Available at: http://segmentnext.com/2017/02/03/super-data-revealed-number-sales-revenue-charts-psvr-rift-gear-vive/ (Accessed: 26 May 2017).

Baker, E. (2017) What Can You Do with 3D Sound That You Can't Do with 2D Sound? No Film School. Available at: http://nofilmschool.com/2017/02/3D-audio-virtual-reality-sound-designer-viktor-phoenix-interview (Accessed: 24 April 2017).

Balazs, B. (2015) Theory of The Film: Character and Growth of a New Art. CreateSpace Independent Publishing Platform.

Doane, M. A. (1985) 'Ideology and the Practice of Sound Editing and Mixing', in Weis, E. and Belton, J. (eds) Film Sound: Theory and Practice. New York: Columbia University Press.

Foder, T. (2016) 'VR Audio: Trends and Challenges of Pioneering a New Field'. Available at: http://designingsound.org/2016/08/vr-audio-trends-and-challenges-of-pioneering-a-new-field/ (Accessed: 7 October 2016).

International Data Corporation (2017) Worldwide Spending on Augmented and Virtual Reality Forecast to Reach $13.9 Billion in 2017, According to IDC, www.idc.com. Available at: http://www.idc.com/getdoc.jsp?containerId=prUS42331217 (Accessed: 21 May 2017).

International Society for Presence Research (2009) Presence defined, International Society for Presence Research. Available at: https://ispr.info/about-presence-2/about-presence/ (Accessed: 7 October 2016).

Jastrow, M. (2016) 'Bringing VR to the Next Level: Technicolor's Marcie Jastrow Discusses Immersive Content Creation and Distribution – The Future Trust', 21 September. Available at: http://thefuturetrust.technicolor.com/article/the-future-of-storytelling/bringing-vr-to-the-next-level-

technicolors-marcie-jastrow-discusses-immersive-content-creation-and-distribution/ (Accessed: 21 May 2017).

Kuzminski, A. (2016) ‘VR Audio: Trends and Challenges of Pioneering a New Field’, Designing Sound, 9 August. Available at: http://designingsound.org/2016/08/vr-audio-trends-and-challenges-of-pioneering-a-new-field/ (Accessed: 7 October 2016).

Lalwani, M. (2016) For VR to be truly immersive, it needs convincing sound to match, Engadget. Available at: http://www.engadget.com/2016/01/22/vr-needs-3d-audio/ (Accessed: 8 February 2016).

Riviere, A. (2016) Why VR will put Game Audio in the Spotlight, AudioKinetic. Available at: http://blog.audiokinetic.com/why-game-audio-will-put-vr-in-the-spotlight (Accessed: 7 October 2016).

Sonnenschein, D. (2002) Sound Design: The Expressive Power of Music, Voice and Sound Effects in Cinema. Studio City, CA: Michael Wiese Productions.

Walden, J. (2016) ‘VR Audio: Crytek goes to new heights for VR game “The Climb”’, Randi Altman’s Post Perspective, 6 June. Available at: http://postperspective.com/vr-audio-crytek-goes-to-new-heights-for-vr-game-the-climb/ (Accessed: 16 May 2017).

Figures