
Journal of New Music Research

ISSN: 0929-8215 (Print) 1744-5027 (Online) Journal homepage: https://www.tandfonline.com/loi/nnmr20

# Reflections from five years of Sonic Interactions in Virtual Environments workshops

Stefania Serafin, Federico Avanzini, Amalia De Goetzen, Cumhur Erkut, Michele Geronazzo, Francesco Grani, Niels Christian Nilsson & Rolf Nordahl

**To cite this article:** Stefania Serafin, Federico Avanzini, Amalia De Goetzen, Cumhur Erkut, Michele Geronazzo, Francesco Grani, Niels Christian Nilsson & Rolf Nordahl (2020) Reflections from five years of Sonic Interactions in Virtual Environments workshops, Journal of New Music Research, 49:1, 24-34, DOI: 10.1080/09298215.2019.1708413

**To link to this article:** https://doi.org/10.1080/09298215.2019.1708413

Published online: 12 Jan 2020.

JOURNAL OF NEW MUSIC RESEARCH
2020, VOL. 49, NO. 1, 24–34
https://doi.org/10.1080/09298215.2019.1708413

Routledge
Taylor & Francis Group



# Reflections from five years of Sonic Interactions in Virtual Environments workshops

Stefania Serafin[a], Federico Avanzini[b], Amalia De Goetzen[a], Cumhur Erkut [a], Michele Geronazzo [a], Francesco Grani[c], Niels Christian Nilsson[a] and Rolf Nordahl[a]

[a]Aalborg University Copenhagen, Denmark; [b]University of Milano, Italy; [c]HERE Technologies GmbH, Germany

**ABSTRACT**
For the past five years, the authors have been running at the IEEE Virtual Reality Conference a Workshop called Sonic Interactions in Virtual Environments (SIVE). The main goal of the workshop series has been to increase among the virtual reality community awareness of the importance of sonic elements when designing multimodal and immersive virtual environments. Starting from this experience, this paper presents a survey of the main active research topics related to sound in virtual and augmented reality (VR/AR), ranging from basic research in spatial audio rendering and sonic interaction design to applications in interactive environments for training, health, rehabilitation, entertainment, and art. Looking at the different research topics emerging from laboratories worldwide, the paper discusses how different research communities can collaborate and benefit from each other in order to increase sound awareness in VR and AR.

**ARTICLE HISTORY**
Received 1 April 2019
Accepted 5 December 2019

**KEYWORDS**
Sonic interaction; virtual reality; spatial audio rendering

## 1. Introduction

Sonic interaction design (SID) is defined as the study and exploitation of sound as one of the principal channels conveying information, meaning, and aesthetic and emotional qualities in interactive contexts. The definition of sonic interaction design emerged thanks to a successful European Research project implemented as a COST Action (COST IC601)[1] (see Franinović & Serafin, 2013; Rocchesso, 2011), where the relevant areas involved in sonic interaction design, such as product sound design, sonification, evaluation and artistic applications were outlined. Sonic interaction can be considered to be laying at the intersection of interaction design and sound and music computing.

In the virtual and augmented reality communities, the focus on research in topics related to auditory feedback has been rather limited when compared, for example, to the focus placed on visual feedback or even on haptic feedback. However, in different domains, such as film or product sound design, it is well known that sound is a powerful way to communicate meaning and emotion to a scene or a product.

In 2014, the authors of this paper started a series of workshops affiliated with the IEEE Virtual Reality Conference, with the name *Sonic Interactions in Virtual Environments* (SIVE).[2] The main goal of the SIVE workshops was to enhance the awareness of the role of sound in interactive experiences, with a focus on sound for virtual and augmented reality. To this end, the workshops promoted discussion on how research in related fields, such as film sound theory, product sound design, sound and music computing, game sound design, and computer music, can inform designers of virtual reality environments. Moreover, the workshop featured state of the art research on the field of sound for virtual environments. Since its first instalment in 2014, SIVE has accepted work by 109 different researchers, affiliated with 42 institutions from 12 different countries.

To date ,the SIVE proceedings include a total of 38 publications. Regarding the topics covered, the papers published in the SIVE proceedings can be divided into six broad categories. Across all workshops, the research belonging to the category *presence and evaluation* has been the most frequent and it accounts for 36.8% (14/38) of the published papers, followed by work on *interaction in virtual environments* which accounts for 23.7% (9/38) of the papers. Papers on *physics-based models* account for 7.9% (3/38), papers on *binaural rendering* account for 15.8% (6/38), papers on *speaker-based reproduction* account for 5.3% (2/38), and finally 10.5% (4/38) have primarily focused on *applications* related to sonic interaction in virtual environments. Figure 1 shows the

**CONTACT** Stefania Serafin sts@create.aau.dk Aalborg University Copenhagen, Denmark

[1] COST actions are European research projects whose goal is to promote collaborations among institutions and exchange young researchers.
[2] Official website of SIVE workshops: sive.create.aau.dk

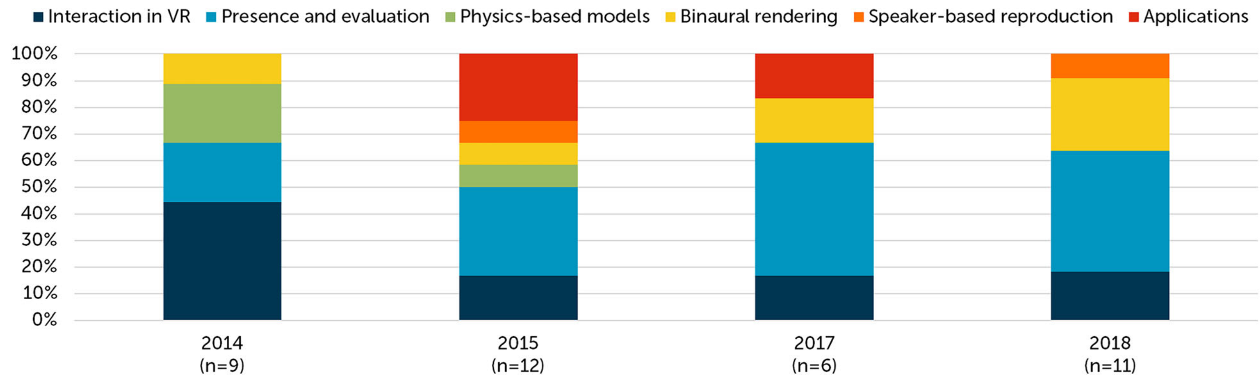


**Figure 1.** Stacked bar chart showing the percentage of papers qualifying as each research type each year SIVE was run. The number of papers included in the proceedings each year is presented along the horizontal axis.

percentage of papers belonging to each of the six categories across the four instalments of SIVE. As apparent from the figure, SIVE has seen an increase in the number of papers covering presence and evaluation, and binaural rendering has also been a regular theme. Finally, it is worth noting that work on virtual reality music instruments (VRMIs) has been featured regularly at SIVE where, 21.1% (8/38) of the papers dealt explicitly with such instruments.

The experience gained through the SIVE workshop editions, and the above analysis of the workshop contributions, constitute the starting point for a wider survey of the main active research topics related to sound in virtual and augmented reality (VR/AR), which is presented in the remainder of this paper. The survey is organised around the six categories discussed above (Sections 2–7) and ranges from basic research to applications in interactive environments for training, health, rehabilitation, entertainment, and art. Looking at the different research topics emerging from laboratories worldwide, we aim at suggesting new ways in which different research communities can collaborate and benefit from each other in order to increase sound awareness in VR and AR.

## 2. Interaction in VR

One of the emerging topics is interaction in immersive environments supporting multiple modalities. Currently, the leading VR companies are continuously developing hand-held controller options at low price points. Such controllers are perhaps the easiest way to interact with a majority of fully immersive VR experiences. Manufacturers and developers create their own interaction designbest-practices[3] and guidelines[4] for a satisfactory and acceptable interaction fidelity (Jerald, 2015). These may explain the sudden decrease in the number of interaction-focused contributions to SIVE from 2014 to 2015, and beyond.

Jerald (2015) has categorised over a hundred VR-based interaction themes into interaction patterns. The resulting 16 patterns were organised into five overarching groups: Selection, Manipulation, Viewpoint Control, Indirect Control, and Compound Patterns. These groups roughly correspond to the categories derived after an extensive task analysis (Jankowski & Hachet, 2015). Also design space approaches, such as the one presented by Blom and Beckhaus (2013) relate to one or several groups.

But an important difference between task or design-space-based approaches and the pattern approach is that the patterns are derived from the user point of view. Jerald motivates his approach by revisiting general interaction models, such as Don Norman's principles and stages of interaction[5], and VR-specific concepts, such as interaction fidelity, proprioceptive and egocentric interaction, reference frames and ergonomic concepts such as cybersickness and fatigue. These patterns can be further broken down into more specific interaction techniques. For example, the Walking Pattern, a form of viewpoint control, consists of real-world walking, redirected walking, walking in place, treadmill walking, etc. The walking pattern is the most popular one in SIVE. It occurs in 13 papers (62 instances), whereas cybersickness is only mentioned in a single paper.

More recently, Gillies (2019) distinguished three broad categories of interaction: object focused, direct mapping, and movement focused. Direct mapping is defined as a form of interaction in which the movements of a user are directly mapped into some form of digital space. A large proportion of Virtual Reality Interaction in which the user's body mapped into the VR

[3] See e.g. https://developer.oculus.com/design/latest/concepts/book-bp/
[4] http://www.uxofvr.com
[5] Mentioned only in one SIVE paper

space is thought to belong to direct-mapping interaction Gillies (2019). In contrast, object-focused interaction design considers the (digital) objects that users can interact with using whatever movements they choose (or at least those that are possible to track). Movement-focused interactions extract the way a movement is performed in immersive environments, shifting the attention from objects and mapping onto the movement itself (Gillies, 2019). Movement-focused interactions are very popular in SIVE: 26 papers mention movement (119 times total), and half of them directly focus on movement.

An example of direct-mapping interaction is a virtual button that is 'clicked' by reaching out and touching it. The interaction is not determined by a specific movement by the user but simply the location of their hand that is mapped into virtual space. Unlike object-focused interaction, the focus is not on the affordances of the button object; they are simply thought to transfer from the real to the virtual world. A key concept here is the affordance, which exploits specific knowledge that users already have of other domains in terms of interaction metaphor (Madsen, 1994). Metaphors were mentioned in seven SIVE papers, whereas affordances appear in six papers. In sonic terms, physics-based sound synthesis models relate to interaction metaphors and affordances the most. Therefore we continue with an overview of physics-based sound synthesis and interaction models.

## 3. Physics-based models

Physics-based sound modelling refers to a set of synthesis algorithms that are based on a description of the physical phenomena involved in sound generation, whereas earlier techniques are based on a description of the sound signal (e.g. in terms of its waveform or its spectrum) and make no assumptions on the sound generation mechanisms.

Researchers in VR and CG have demonstrated a growing interest in the use of these techniques. During the last decade, several related contributions have appeared in the proceedings of major international conferences such as IEEE VR and Siggraph, and courses have been presented to the scientific community (James, 2016). This interest is partly reflected also in SIVE contributions, although to a lesser extent (Baldan, Lachambre, Delle Monache, & Boussard, 2015; Rausch, Hentschel, & Kuhlen, 2014).

Since physically based models generate sound from computational structures that respond to physical input parameters, they automatically incorporate complex responsive acoustic behaviours. A second advantage is interactivity and ease in associating motion to sound control. As an example, the parameters needed to characterise impact sounds (e.g. relative normal velocity), are computed in a VR physical simulation engine and can be directly mapped into control parameters, producing a natural response of the auditory feedback to user gestures and actions. Finally, physically based sound models can in principle allow the creation of dynamic virtual environments in which sound-rendering attributes are incorporated into data structures that provide a multimodal encoding of object properties (shape, material, elasticity, texture, mass, etc.).

Following the categorisation proposed by Gaver (1993) in the framework of ecological acoustics, the sound sources modelled by physics-based approaches may be clustered into three main categories: solids, liquids, gases.

### 3.1. Solids

Vibrations in solids can be synthesised through numerical simulations based on finite-difference and finite-element models (Bilbao, 2009; Doutaut, Matignon, & Chaigne, 1998; Lambourg, Chaigne, & Matignon, 2001; O'Brien, Shen, & Gatchalian, 2002): finite-element simulations are employed for the generation of both animated video and audio. Complex audio-visual scenes can be simulated, but heavy computational loads still prevent real-time rendering and the use of these methods in interactive applications.

A more efficient technique is modal sound synthesis originally proposed by Adrien (1991). Starting with the studies by Van den Doel and coworkers (van den Doel, Kry, & Pai, 2001; van den Doel & Pai, 1998), this has become the most used approach for the simulation of non-musical sounds produced by mechanical contact of solid objects. Modal representations can be derived both for discrete and continuous oscillating systems. In both cases, the vibration of an object at a given point can be seen as a linear combination of modal displacements, which in turn are described via a set of uncoupled second-order oscillator equations. As an example, a string (continuous system) can be approximated with the discrete network of Figure 2, with $N$ masses. The discrete system has $N$ modes, whose shapes resemble more and more closely those of the continuous system, as $N$ increases.

Despite the comparatively low computational costs with respect to other techniques, mode-based numerical schemes can become expensive when many objects, each with many modes, are rendered simultaneously. Several studies have dealt with optimising modal synthesis schemes and using perceptual criteria to perform mode compression and truncation, so as to reduce

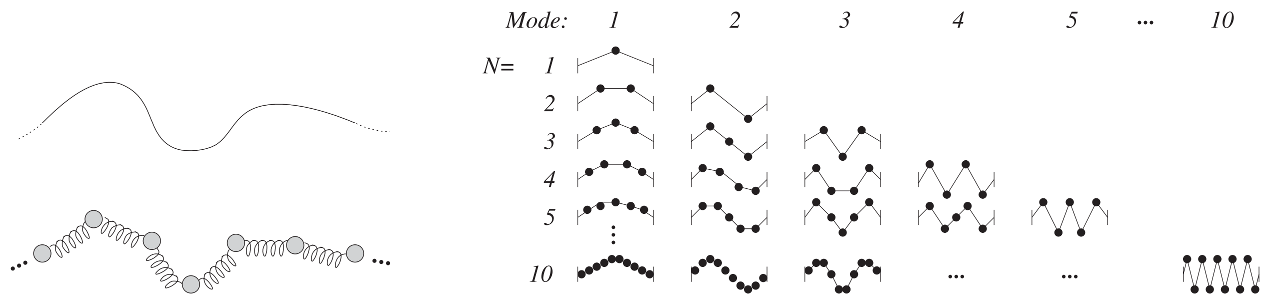


**Figure 2.** Modal descriptions of continuous and discrete systems. Left: approximation of an ideal string with a mass-spring network. Right: modes of the discrete system for various numbers *N* of masses.

the computational load (Lloyd, Raghuvanshi, & Govindaraju, 2011; Raghuvanshi & Lin, 2007). The use of variable time-steps in the integration of the numerical equation has also been explored (Zheng & James, 2011). Recent works have exploited parallelisation, such as Advanced Vector Extensions (AVX) to the x86 instruction set architecture (van Walstijn & Mehes, 2017).

With specific regard to SIVE contributions, Rausch et al. (2014) presented various algorithms to distribute the computations for modal sound synthesis on a GPU and compared their results to CPU implementations. The authors were able to show that the GPU algorithms provide a significantly higher performance, and in particular, allow to synthesise a large number of sounding objects simultaneously.

One further current research direction concerns improvements to the basic modal synthesis scheme, with the aim of increasing the realism and quality of the resulting sounds. One of the key challenges is the development of automatic estimation of modal parameters that recreate realistic audio. Typical approaches are based on the analysis of pre-recorded target audio clips (Ren, Yeh, & Lin, 2013), or on offline numerical simulations (Picard, Frisson, Faure, Drettakis, & Kry, 2010). Estimating the damping parameters of materials is especially problematic, as external factors can interfere with accurate estimation (Sterling, Rewkowski, Klatzky, & Lin, 2019). Finally, modal amplitude variations and spatialisation effects due to acoustic wave radiation also need to be accounted for. Precomputed Acoustic Transfer methods allow radiation fields to be precomputed by numerical solvers and then approximated by efficient representations that can be evaluated for real-time rendering (James, Barbič, & Pai, 2006; Langlois, An, Jin, & James, 2014; Li, Fei, & Zheng, 2015).

### 3.2. Liquids and gases

Sounds produced by aerodynamic mechanisms and by liquids are also actively studied, although to a lesser extent (possibly because of their inherent reduced interactivity with respect to solid interactions).

Given the great variety of possible liquid sounds (ranging from stochastic sounds such as that of streaming river, to deterministic ones such as dripping), their synthesis remains a complicated task. Existing research has focused on simulating some of the specific mechanisms responsible for sound generation in liquids, particularly bubble formation. After being formed in a liquid, a bubble emits a decaying sinusoidal sound. If bubble formation occurs close enough to the liquid–air interface, the pitch rises as it approaches the surface. The physical mechanism responsible for these sounds is the pulsation of the bubble volume (Minnaert, 1933): any bubble being a small compressible air region surrounded by incompressible fluid, it oscillates like a spring amid a liquid domain.

Various studies have dealt with bubble sound synthesis, starting from the seminal work by van den Doel (2005), which allowed simulation of more complex liquid sounds (from dripping to heavy rain or waterfalls) through the synthesis of a large population of bubbles. Zheng and James (2009) proposed a similar approach to acoustic bubble simulation, with the aim of augmenting existing numerical solvers for incompressible liquid simulations that are commonly adopted in the computer graphics literature. Examples for various liquid sounds were proposed (including pouring, babbling, and splashing phenomena). Moss, Yeh, Hong, Lin, and Manocha (2010) also proposed a simplified, physically inspired model for bubble creation, designed specifically for real-time applications. Drioli and Rocchesso (2012) proposed a multi-rate approach to the sound synthesis of liquid phenomena, in which smoothed particle motion simulated at low-rate is used to model liquids in motion and to control audio-rate sound synthesis algorithms of basic acoustic events.

Aeroacoustic sounds are also extremely varied and include cavity tones, edge tones (such as those produced by organ pipes), aeolian tones (produced e.g. by swinging objects), hole tones, and turbulence noises (jet engines, explosions, and so on). Typical approaches to the synthesis of such sounds may be defined as being 'physically inspired' rather than truly physics based. This definition means that these are hybrid approaches that replicate the signal produced but add characteristics of the physics that are behind the sound creation. As an example, the sound of a swinging sword may be modelled as noise shaping with a bandpass filter with centre frequency proportional to the speed of the swing. A recent paper on the synthesis of swinging objects also provides an exhaustive analysis of the state of the art in this field (Selfridge, Moffat, & Reiss, 2017).

Within the ecological taxonomy proposed by Gaver (1993), engine sounds represent a very specific category of aeroacoustic sounds, and yet a very relevant one in terms of potential applications: the engine sound perceived in the cabin conveys relevant cues about the vehicle motion. These sounds have been the subject of one SIVE contribution (Baldan et al., 2015). The authors presented a procedural and physically informed model for synthetic combustion engine sound, which provides an effective and flexible tool for designing and simulating engine sounds and provides control over several parameters (two-/four-stroke cycle, number of cylinders, etc.). The model was implemented in real-time and was integrated on a driving simulator environment for industrial sound design.

## 4. Binaural sound rendering

A high fidelity but efficient sound simulation is an essential element in immersive virtual reality (VR) and SIVE community. Continuous advances in hardware and software technologies foster interaction between virtual sounds and humans in rendering experiences with an increasing level of realism and presence. On the other hand, a perceptually plausible and efficient auralisation is forced to be preferred to an authentic rendering due to the limitation in memory and computational power related to low latency constraints. This trade-off is complex and challenging because, especially in VR, real-time constraints involve a multimodal system, thus requiring resources shared with graphics, sensors, application logic, and high-level functionality (e.g. artificial intelligence).

From the literature (Savioja, Huopaniemi, Lokki, & Väänänen, 1999), one can refer to the general term *auralisation* that covers the three main components for sound spatialisation considered in some of the SIVE contributions: (i) receiver modelling (see Section 4.1), (ii) room acoustics modelling (see Section 4.2), and (iii) source modelling (see Section 3).

### 4.1. Headphone reproduction with head-related transfer functions

One of the main advantages of headphone-based reproduction is the complete control of sound synthesis and binaural cues arriving at each ear (left and right channels); high level of isolation from external and noisy environmental sounds, e.g. echoes and reverberation, in listeners' can be obtained by ear occlusion and noise-cancelling technologies. However, headphones may be experienced as intrusive by the user at the expense of naturalness and externalisation of the listening experience (Hale & Stanney, 2002). Headphone-induced spectral colouration can be reduced by carefully following product design criteria and ad-hoc equalisation algorithms with the aim of minimising artefacts in the binaural audio reproduction.

Auralisation with headphones requires head-related transfer function (HRTF) that encodes spatial-temporal acoustic properties of a human body resulting from the interaction of user's head, ear, and torso with the soundfield in space. The synthesis of a binaural anechoic spatial sound can be obtained by the convolution of an anechoic sound signal with left- and right-ear HRTFs chosen among a discrete set of measurements, i.e. spatial locations, and interpolation in a suitable functional basis, like in the spherical harmonic (SH) domain. A HRTF spatial grid of 4–5° spacing in both azimuth and elevation with decreasing density above the head leads to a perceptually optimal representation for HRTF filters that could be convolved with partitioned block algorithm providing a compromise between computational efficiency and latency (Välimäki, Parker, Savioja, Smith, & Abel, 2016).

In multimodal virtual environments, a common approach is to use the same *generic* HRTFs, e.g. recorded using a dummy head (see Figure 3 for some examples), for any listener in order to obtain a trade-off between average efficacy and measurement/personalisation procedures taking into account the dominance of visual cues for localisation. However, generic HRTFs generally introduce unnatural colouration in the frequency spectrum and degradations in localisation and immersion of the listening experience. Paul (2009) provided a historical review on this topic.

Recent literature is increasingly investigating the development of *personalised* HRTFs for each individual user in virtual/augmented reality in order to support a listening experience which is perceptually equivalent to that with own individual HRTFs. HRTFs can be computed from basic geometry of the body and/or accurate numerical simulations with boundary element (BEM) and finite-difference time-domain (FDTD) methods (Prepelita, Geronazzo, Avanzini, & Savioja, 2016;

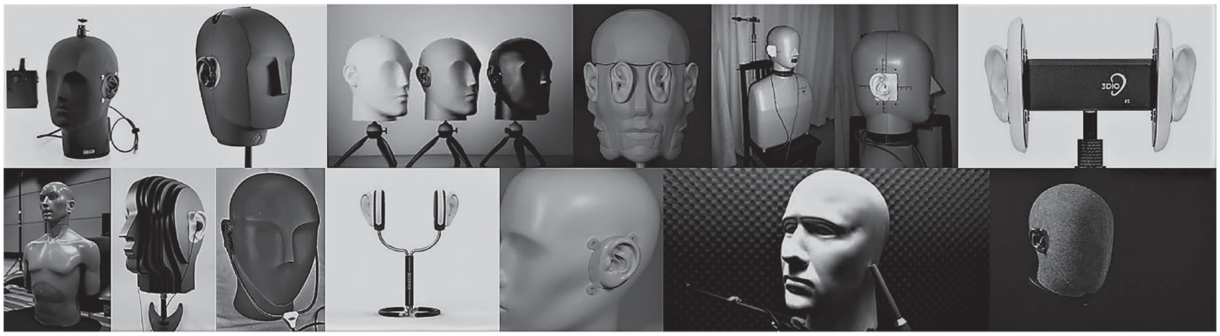

**Figure 3.** Examples of dummy-heads microphones for binaural recording and reproduction.

Xie, 2013). Moreover, a structural interpretation of the acoustic contribution of head, pinna, shoulders and torso can guide filter modelling of time and spectral (e.g. peaks and notches) variations (Geronazzo, Spagnol, & Avanzini, 2013). HRTF features can be also defined according to principal component analysis (PCA) and subsequently personalised with self-tuning actions of weights (Hwang, Park, & Park, 2010). On the other hand, optimised selection procedures of existing non-individual HRTFs is an alternative approach in which HRTFs are chosen by selecting the best match among several HRTF sets. These approaches can benefit from the exponential increase of available HRTF data during the last 10 years, which has also provided motivations for standardisation processes such as the Spatially Oriented Format for Acoustics (SOFA) (Majdak et al., 2013).[6]

SIVE workshops considered contributions to the definition of standards for HRTF measurement and simulation setups (Barumerli, Geronazzo, & Avanzini, 2018) and the development of research frameworks for systematic HRTF evaluations with computation auditory models (Geronazzo, Carraro, & Avanzini, 2015) in order to provide insights regarding key aspects in robust assessments of future auralisation technologies.

### 4.2. Room acoustic modelling

The common and simple approach to provide an approximation of an acoustical space is by using a static RIR which could be convolved with an original dry signal; unfortunately, this method lacks flexibility and it is inappropriate for VR. Rendering sound propagation for VR requires the spatialisation of the directional RIR in the *spatial room impulse response* (SRIR).

Interactive auralisation forces algorithms to cover most of the psychoacoustic effects in localisation and timbre distortion due to dynamic changes of the active listening experience, thus defining memory and computational requirements. Given an encoding representation of the sound field, interactive VR latency requires the computation of SRIR in a convenient way: for instance, Schissler, Nicholls, and Mehra (2016) performed the convolution in the spherical harmonics domain for HRTFs, sharing similar aspects to technologies for multi-channel surround systems, such as high-order ambisonics, wave-field synthesis, and directional audio coding.

Sound propagation simulates the acoustics of the space, either a closed space such as a room or an open space surrounding the listeners (a complete survey in interactive virtual environments is provided in Välimäki et al., 2016). One of the main challenges is the accurate modelling of sound propagation which is computationally intensive due to the multitude of reverberation paths from virtual sound sources to listener ears/listening area.

Perceptually motivated algorithms provide the control of computational resources, i.e. CPU and GPU processing, for SRIR computation, allowing a flexible scaling of aural complexity. For this reason, algorithms for dynamic spatialisation of room acoustics allow immersion and externalisation of a multiple sound source scenarios with configurable geometric complexity and implemented acoustic effects (i.e. order of reflections, diffraction, and scattering). Diffracted occlusion and sound propagation of early-reflections should be coherently rendered in order to implement a perceptually plausible auralisation.

Sound propagation modelling can be classified into three main approaches:

- **geometric methods**, involving a high-frequency approximation of sound propagating in the form of rays;
- **wave-based methods**, solving underlying physical equations;
- **hybrid methods**, mixture of the previous approaches.

In one of the SIVE contributions, Mehra and Manocha (2014) described an interactive sound propagation system based on the equivalent source method (ESM) for

[6] See also the official website of the SOFA project http://sofaconventions.org/ for an exhaustive list of freely available HRTF databases).

realistic outdoor sounds. This method extends previous researches where pressure fields are pre-computing based on elementary spherical harmonic (SH) description of the wave-based sound propagation in the frequency domain; the high-dimensional acoustic field is compressing and allocated in memory for dynamic changing of sources/listener position and source directivity.

## 5. Speaker based reproduction

On loudspeaker-based spatial sound reproduction, music composers and engineers have been pushing the boundaries of what is technically feasible since decades, tracing back the first attempts of electroacoustic spatial sound reproduction on arrays of loudspeakers to the middle of the twentieth century. The range of techniques currently available to create virtual surround environments with loudspeakers is broad and spans across various multi-channel configurations, from quadraphonic and ITU 5.1 Surround (International Telecommunication Union, 2012), to more extended techniques such as, e.g. VBAP (Pulkki, 1997), DBAP (Lossius & Baltazar, 2009), Ambisonics (Gerzon, 1985), wave-field synthesis (WFS) (Berkhout, 1988; Berkhout, de Vries, & Vogel, 1993), and more. Although for the individual consumption of virtual experiences the delivery of soundscape through headphones presents many advantages as discussed in Section 4.1, SIVE has also demonstrated use cases in which loudspeaker-based sound reproduction serves as the preferred technique.

Collaborative music experiences represent a typical case in which the fruition – or the collaborative creation – of the experience by multiple users at the same time is better enabled by the reproduction of sound through arrays of loudspeakers. In this case, users are free to roam around the physical space, keep communicating with each other without any impediment possibly caused by wearing headphones, and to experience the natural acoustics of the physical space in which they are located together with the sound content of the virtual experience.

An example of a collaborative virtual experience is described in Pfaff and Lervik (2015). In this work, the authors presented a virtual environment where different game mechanics function as the playground for collaborative music making. Scope of their work being to investigate whether introducing game elements and the possibility to freely create music within a virtual environment would let emerge new musical and social patterns. Two possible target versions of the game were made. One intended for home use, in which single players are collaboratively creating the music experience in remote connection with each other. Another one, referred to as 'Concert Version', in which up to six users/players take part in the experience. For the Concert Version, the players are located within a surround installation of eight loudspeakers rendering an Ambisonics Gerzon (1985) soundfield; part of the game elements offered to the players is the control of displacement and reverberation of sound sources within the virtual space. As the outcome of their research, authors argue how this type of collaborative music making experiences might lead to an entirely new way of perceiving music.

Within the same topic of virtual experiences in which sound plays the dominant role and where by intent the experience is meant to be a collective one for multiple users, there is the experiment of Grani, Nordahl, and Serafin (2016); Lind et al. (2017). In this case, the purpose of reproducing the real feeling of presence given by the experience of standing in the middle of a crowd at a live rock concert is re-enacted by coupling the 360 video recording of a live show with a spatial sound rendering of the sound mix delivered via Wavefield Synthesis. The authors of this study aim to compensate the lack of social presence (i.e. the lack of experiencing the concert with friends and other audiences) typically given by watching a concert in VR, with the introduction of focussed sound sources reproducing audience noise within the listening space. Although a comparative study did not demonstrate the significative impact of the addition of such additional virtual sources on the perception of presence, qualitative results show that the naturalness of the sonic experience delivered through wavefield synthesis had a positive impact on the participants.

For music composers, though, the main obstacle to multi-channel sound spatialisation is still represented by the difficulty of handling a multi-speaker set-up as a complex instrument in itself. An attempt to ease the onboarding process for composers has been presented at SIVE in the work of Timmermans (2015). A multi-channel set up is seen as a complex instrument on which a composition could be performed and for which a composition could be created. The author of this work proposes a combination of pre-coded presets to implement sound trajectories, and interpolation techniques to automate the processes of panning and spatial mix of trajectories. The tool proposed aims to be integrated within the composition workflow, in a c-like scripting language in which both sounds, spatial trajectories and musical structures are coded. Each sound in a mix is an object consisting of individual position, trajectory, envelope and acoustic properties, e.g. characteristics of reverberation. Intuitive naming of variables is proposed in a sort of markup language where 'frontLeft', 'crossLeft' and similar names can be given on-the-go by the composer during the creative process of musical composition, leaving the actual coding for a later stage. Novelty of the proposed approach is the consideration of a sound as an ”object” that includes all its acoustic properties including spatial characteristics and described into the music score as a preset on a time-based script, leaving the composer out from the traditional track-bus-mixer DAW approach. Given also in this case the collective nature of the experience (electroacoustic concerts) for which the tool was presented at SIVE, loudspeaker-based reproduction represents a promising way to deliver non-individual virtual experiences.

## 6. Presence and evaluation

The concept of presence has several different definition and interpretations, especially in the music engineering community. In the VR community, presence can be defined as the sensation of being in a virtual environment. Presence has received increasing attention in the last decades, and this is also reflected in the SIVE papers. However, despite the growing attention in investigating presence, the influence of the auditory modality remains relatively unexplored, compared with its visual counterpart.

One of the prominent researchers in the field of presence research is Mel Slater. In his articles, he distinguishes between immersion and presence, where immersion is given by the technical capabilities of a system (e.g. in the case of sound, the number of speakers in a surround sound system), while presence refers to the human's reaction. He particularly identifies two elements of presence, named place illusion and plausibility illusion (Slater, 2009). Place illusion is defined as the sensation of being in a specific place, while plausibility illusion is defined as the extent to which the system can produce events that directly relate to the participant, and the overall credibility of the scenario being depicted in comparison with expectations. In (Nordahl & Nilsson, 2014) the authors present an overview of the research performed in the field of presence and VR with a specific focus on the sound of being in a specific space.

In the context of SIVE, presence has been addressed from several perspectives, ranging from the role of a performer of a virtual reality musical instrument, to the evaluation of the quality of a sound delivery system, to the role of self-sounds in enhancing presence. In Berthaut, Zappi, and Mazzanti (2014), the authors distinguish between presence of the performer versus presence of the audience exposed to an immersive scenography.

An interesting topic of research is the synthesis and perception of self-sound, and how this affects presence in VR. As an example, one SIVE contribution proposes an experiment comparing three presentation formats (audio only, video with audio and an interactive immersive VR format) and their influences on a sound design evaluation task concerning footstep sounds. The evaluation involved estimating the perceived weight of a virtual avatar seen from a first person perspective, as well as the suitability of the sound effect relative to the context. The results show significant differences for three cases between the presentation formats, both for weight estimates and suitability ratings over all variations of the footstep sound design (Sikstrom, De Gotzen, & Serafin, 2015). One of the recent SIVE contributions examines this topic further, with a focus on presence and both footstep sounds and environmental sounds (Kern & Ellermeier, 2018). In this paper, subjects were exposed to $2 \times 2$ conditions (with and without footsteps, and with and without environmental sounds), and presence was measured using the IPQ presence questionnaire (Schubert, Friedmann, & Regenbrecht, 2001). Results show that self-generated footsteps had a significant impact on the sound of being there and perceived realism, and adding a soundscape enhanced the effect. Overall SIVE workshops have contributed to providing a stronger understanding on the role of sound in enhancing presence and immersion in VR.

## 7. Applications

In the last decade VR has been proved to be beneficial in many different fields such as entertainment, education, art and health, just to name the most important ones. The development of the technology in each of these areas has seen audio as an ancillary feature until few years ago. The SIVE workshop has helped the VR community to recognise the importance of sound to create more immersive and effective experiences and the Sound and Music research community to discover VR for both artistic and educational purposes. Immersive Virtual Musical Instruments (IVMIs) are for instance a new means to explore from inside old and new musical instruments that can also be played in front of an audience as part of a scenography. In his paper, Berthaut et al. (2014) discuss different IVIMIs stage setups to showcase live virtual performances and their scenographic level. VR though can also easily allow a composer to interact with more than one instrument up to a full virtual orchestra where the instruments can be of very peculiar nature, as in the case for the virtual orchestra of factory machinery (Simon, Nouviale, Gaugne, & Gouranton, 2014) where two users with a remote controller can trigger and control punctual sounds, animations, sound and visual loops, as well as the global tempo of the application. Collaboration between users within the same VR environment is a particularly interesting topic in the context of collaborative music making (Men & Bryan-Kinns, 2018;

Pfaff & Lervik, 2015). In the SIVE workshops, collaborative music making has been approached with two different focuses. The LeMo system (Men & Bryan-Kinns, 2018) was developed to see if VR could provide a new meaningful way of making 3D annotations to support the composition process and the human to human interaction process, while gamification and game mechanics applied to collaborative music making was at the core of the work by Pfaff and Lervik (2015). In both cases, the authors explored new ways of creating and perceiving music in VR.

In the field of education and training, there are countless VR applications, that spans from simulators of various nature where drivers, soldiers or surgeons can train their skills according to the nature of the simulation, to applications for kids, where they can visit historical locations or simply learn different disciplines through an immersive and interactive experience. In the SIVE workshops VR has been used to explore and represent music compositional structures and harmonic relationships (Mandanici, Roda, & Canazza, 2015) as well as to investigate modulations in motor preparation timing induced by sounds, allowing a direct experience of the close connection between sound and movement (Geronazzo, Nardello, & Cesari, 2018). Training navigation skills of blind children is the aim of another interesting application that has been presented at the SIVE workshop: the Legend of Iris (Allain et al., 2015), a 3D navigation game that generates an accurate and realistic soundscape that is used by the player to navigate in the virtual environment. Sonic interaction design and sound spatialisation techniques can be in fact very useful to improve the user's experience of a VR game as discussed in different SIVE papers (Mehra, Rungta, Golas, Lin, & Manocha, 2015; Summers & Jesse, 2017; Summers, Lympouridis, & Erkut, 2015). As a last area of application, it is worth mentioning health and well being: new studies are now exploring how to use VR to help autistic children to deal with their everyday challenges and to reduce stress (Adjorlu & Serafin, 2019). In general, VR can provide a very safe and controlled environment where to explore different sources of anxiety and possibly overcome them. In the context of sound and music computing, it can be also particularly relevant to be able to simulate hearing loss of different severity to allow a higher degree of understanding between the patient and the doctor that many times will have to tune the hearing aid together (Cuevas-Rodriguez et al., 2018). To summarise, the SIVE workshops have been a great arena of discussion of possible VR applications and also a privileged point of observation of the growing VR community, more and more conscious of the value of SID.

## 8. Conclusions

The SIVE workshops are an opportunity for the authors of this paper to be connected with the VR community at large and stress the importance of sound to create compelling and engaging interactive experiences. We noticed that the community working on sound for VR is rather scattered, ranging from those interested merely in sound reproduction systems, and therefore targeting merely venues like Audio Engineering Society conventions, to those interested in VR-based musical instruments, and therefore targeting venues such at the New Interfaces for Musical Expression community or the Sound and Music Computing community. Our interest in creating and sustaining the SIVE series of workshops has been merely to address the need for better research, development and awareness of the role of sound in multimodal immersive environments. We have noticed in several occasions that several VR experiences could have benefitted from more careful sound design. Moreover, the field of procedural based sound synthesis for VR is still at its infancy and can grow thanks to combined research efforts from different disciplines.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

*Cumhur Erkut* http://orcid.org/0000-0003-0750-1919
*Michele Geronazzo* http://orcid.org/0000-0002-0621-2704

## References

Adjorlu, A., & Serafin, S. (2019). Virtual reality (VR) for children diagnosed with autism spectrum disorder (ASD): Interventions to train social and everyday living skills. In G. Guazzaroni (Ed.), *Virtual and augmented reality in mental health treatment* (pp. 159–175). IGI global.

Adrien, J.-M. (1991). The missing link: Modal synthesis. In G. De Poli, A. Piccialli, & C. Roads (Eds.), *Representations of musical signals* (pp. 269–297). Cambridge, MA: MIT Press.

Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R., & Gaubitch, N. D. (2015). An audio game for training navigation skills of blind children. In *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)*, Arles, France (pp. 1–4). IEEE.

Baldan, S., Lachambre, H., Delle Monache, S., & Boussard, P. (2015). Physically informed car engine sound synthesis for virtual and augmented environments. In *Proceedings of IEEE international workshop on sonic interactions for virtual environments (SIVE2014)*, Arles.

Barumerli, R., Geronazzo, M., & Avanzini, F. (2018). Round robin comparison of inter-laboratory hrtf measurements – assessment with an auditory model for elevation. In *2018*

*IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*, Reutlingen, Germany (pp. 1–5).
Berkhout, A. J. (1988). A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, *36*(12), 977–995.
Berkhout, A. J., de Vries, D., & Vogel, P. (1993). Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, *93*(5), 2764–2778.
Berthaut, F., Zappi, V., & Mazzanti, D. (2014). Scenography of immersive virtual musical instruments. In *2014 IEEE VR Workshop: Sonic interaction in virtual environments (SIVE),* Minneapolis, MN (pp. 19–24). IEEE.
Bilbao, S. (2009). *Numerical sound synthesis – finite difference schemes and simulation in musical acoustics*. Chichester:John Wiley & Sons.
Blom, K. J., & Beckhaus, S. (2013). The design space of dynamic interactive virtual environments. *Virtual Reality*, *18*, 101–116.
Cuevas-Rodriguez, M., Picinali, L., Gonzalez-Toledo, D., Garre, C., de La Rubia-Cuestas, E., Molina-Tanco, L., & Poirier-Quinot, D. (2018). The 3d tune-in toolkit – 3d audio spatialiser, hearing loss and hearing aid simulations. In *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*, Reutlingen, Germany. IEEE.
Doutaut, V., Matignon, D., & Chaigne, A. (1998). Numerical simulations of xylophones. II. Time-domain modeling of the resonator and of the radiated sound pressure. *The Journal of the Acoustical Society of America*, *104*(3), 1633–1647.
Drioli, C., & Rocchesso, D. (2012). Acoustic rendering of particle-based simulation of liquids in motion. *Journal on Multimodal User Interfaces*, *5*, 187–195.
Franinović, K., & Serafin, S. (Eds.). (2013). *Sonic interaction design*. Cambridge, MA: MIT Press.
Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, *5*(1), 1–29.
Geronazzo, M., Carraro, A., & Avanzini, F. (2015). Evaluating vertical localization performance of 3d sound rendering models with a perceptual metric. In *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)* (pp. 1–5). Arles: IEEE Computer Society.
Geronazzo, M., Nardello, F., & Cesari, P. (2018). An educational experience with motor planning and sound semantics in virtual audio reality. In *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*, Reutlingen. IEEE.
Geronazzo, M., Spagnol, S., & Avanzini, F. (2013). Mixed structural modeling of head-related transfer functions for customized binaural audio delivery. In *Proceedings of the 18th international Conference on digital signal process*. (DSP 2013) (pp. 1–8). Santorini, Greece.
Gerzon, M. A. (1985). Ambisonics in multichannel broadcasting and video. *Journal of the Audio Engineering Society*, *33*(11), 859–871.
Gillies, M. (2019). Understanding the role of interactive machine learning in movement interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *26*, 5.
Grani, F., Nordahl, R., & Serafin, S. (2016). Multimodal interactions, virtual reality and 360 movies: Applications using wavefield synthesis. In *Proceedings of the audio mostly 2016*, Piteå, Sweden (pp. 20–27). ACM.
Hale, K. S., & Stanney, K. M. (Eds.). (2002). *Handbook of virtual environments: Design, implementation, and applications*. Boca Raton: CRC Press.
Hwang, S., Park, Y., & Park, Y.-s. (2010). Customization of spatially continuous head-related impulse responses in the median plane. *Acta Acustica United with Acustica*, *96*(2), 351–363.
International Telecommunication Union. 2012. *Multichannel stereophonic sound system with and without accompanying picture*. Electronic Publication Geneva. Retrieved from https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-3-201208-I!!PDF-E.pdf
James, D. L. (2016). Physically based sound for computer animation and virtual environments. In *ACM SIGGRAPH 2016 courses*, SIGGRAPH '16 (pp. 22:1–22:8). Anaheim, CA.
James, D. L., Barbič, J., & Pai, D. K. (2006). Precomputed acoustic transfer: Output-sensitive, accurate sound generation for geometrically complex vibration sources. *ACM Trans. Graphics*, *25*(3), 987–995.
Jankowski, J., & Hachet, M. (2015). Advances in interaction with 3D environments. *Computer Graphics Forum*, *34*, 152–190.
Jerald, J. (2015). *The VR book*. New York, NY: ACM.
Kern, A. C., & Ellermeier, W. (2018). Influence of hearing your steps and environmental sounds in VR while walking. In *2018 IEEE 4th VR workshop on sonic interactions for virtual environments* (SIVE), Reutlingen (pp. 1–4). IEEE.
Lambourg, C., Chaigne, A., & Matignon, D. (2001). Time–domain simulation of damped impacted plates: II. Numerical models and results. *The Journal of the Acoustical Society of America*, *109*(4), 1433–1447.
Langlois, T. R., An, S. S., Jin, K. K., & James, D. L. (2014). Eigenmode compression for modal sound models. *ACM Trans. Graphics*, *33*(4), 40:1–40:9.
Li, D., Fei, Y., & Zheng, C. (2015). Interactive acoustic transfer approximation for modal sound. *ACM Trans. Graphics*, *35*(1), 2:1–2:16.
Lind, R. B., Milesen, V., Smed, D. M., Vinkel, S. P., Grani, F., Nilsson, N. C., & Reng, L. (2017). Sound design in virtual reality concert experiences using a wave field synthesis approach. In *2017 IEEE virtual reality (VR)*, Los Angeles, CA (pp. 363–364). IEEE.
Lloyd, B., Raghuvanshi, N., & Govindaraju, N. K. (2011). Sound synthesis for impact sounds in video games. In *Proceedings of the ACM symposium on interactive 3D graphics and games (ACM I3D)* (pp. 55–62). San Francisco, CA.
Lossius, T., & Baltazar, P. (2009). Dbap–distance-based amplitude panning. In *International Computer Music Conference (ICMC)*, Montreal, Canada.
Madsen, K. H. (1994). A guide to metaphorical design. *Communications of the ACM*, *37*, 57–62.
Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., & Suzuki, Y. (2013). Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions. In *Audio engineering society convention 134*, Rome, Italy. Audio Engineering Society.
Mandanici, M., Roda, A., & Canazza, S. (2015). A conceptual framework for motion based music applications. In *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)*, Arles, France (pp. 1–5). IEEE.
Mehra, R., & Manocha, D. (2014). Wave-based sound propagation for VR applications. In *2014 IEEE VR workshop:*

*Sonic interaction in virtual environments (SIVE)*, Minneapolis, MN (pp. 41–46). IEEE.
Mehra, R., Rungta, A., Golas, A., Lin, M., & Manocha, D. (2015). WAVE: Interactive wave-based sound propagation for virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, *21*(4), 434–442.
Men, L., & Bryan-Kinns, N. (2018). Lemo: Supporting collaborative music making in virtual reality. In *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*. IEEE.
Minnaert, M. (1933). On musical air-bubbles and the sounds of running water. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *16*, 235–248.
Moss, W., Yeh, H., Hong, J.-M., Lin, M. C., & Manocha, D. (2010). Sounding liquids: Automatic sound synthesis from fluid simulation. *ACM Transactions on Graphics*, *29*, 21.
Nordahl, R., & Nilsson, N. C. (2014). The sound of being there: Presence and interactive audio in immersive virtual reality. In K. Collins, B. Kapralos, & H. Tessler, (Eds.), *Oxford handbook of interactive audio*. New York, NY: Oxford University Press.
O'Brien, J. F., Shen, C., & Gatchalian, C. M. (2002). Synthesizing sounds from rigid-body simulations. In *Proceedings of the ACM SIGGRAPH 2002* (pp. 175–181). San Antonio, TX.
Paul, S. (2009). Binaural recording technology: A historical review and possible future developments. *Acta Acustica United with Acustica*, *95*(5), 767–788.
Pfaff, S., & Lervik, O. (2015). Games in concert. In *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)*, Arles, France (pp. 1–4). IEEE.
Picard, C., Frisson, C., Faure, F., Drettakis, G., & Kry, P. (2010). Advances in modal analysis using a robust and multiscale method. *EURASIP Journal on Advances in Signal Processing*, *2010*, 392782.
Prepelita, S., Geronazzo, M., Avanzini, F., & Savioja, L. (2016). Influence of voxelization on finite difference time domain simulations of head-related transfer functions. *The Journal of the Acoustical Society of America*, *139*(5), 2489–2504.
Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, *45*(6), 456–466.
Raghuvanshi, N., & Lin, M. C. (2007). Physically based sound synthesis for large-scale virtual environments. *IEEE Computer Graphics and Applications*, *27*(1), 14–18.
Rausch, D., Hentschel, B., & Kuhlen, T. (2014). Efficient modal sound synthesis on GPUs. In *Proceedings of the IEEE international workshop on sonic interactions for virtual environments (SIVE2014)* (pp. 13–18). Minneapolis, MN.
Ren, Z., Yeh, H., & Lin, M. C. (2013). Example-guided physically based modal sound synthesis. *ACM Trans. Graphics*, *32*(1), 1.
Rocchesso, D. (2011). *Explorations in sonic interaction design*. Berlin: Logos Verlag.
Savioja, L., Huopaniemi, J., Lokki, T., & Väänänen, R. (1999). Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, *47*(9), 675–705.
Schissler, C., Nicholls, A., & Mehra, R. (2016). Efficient HRTF-based spatial audio for area and volumetric sources. *IEEE Transactions on Visualization and Computer Graphics*, *22*(4), 1356–1366.
Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266–281.
Selfridge, R., Moffat, D., & Reiss, J. D. (2017). Sound synthesis of objects swinging through air using physical models. *Appl. Sci.*, *7*(11), 1–18. doi:10.3390/app711117.
Sikstrom, E., De Gotzen, A., & Serafin, S. (2015). Avatar weight estimates based on footstep sounds in three presentation formats. In *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)*, Arles, France (pp. 1–6). IEEE.
Simon, L. S., Nouviale, F., Gaugne, R., & Gouranton, V. (2014). Sonic interaction with a virtual orchestra of factory machinery. In *2014 IEEE VR workshop: Sonic interaction in virtual environments (SIVE)*, Minneapolis, MN (pp. 7–12). IEEE.
Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557.
Sterling, A., Rewkowski, N., Klatzky, R. L., & Lin, M. C. (2019). Audio-material reconstruction for virtualized reality using a probabilistic damping model. *IEEE Transactions on Visualization and Computer Graphics*, *25*(5), 1855–1864. doi:10.1109/tvcg.2019.2898822.
Summers, C., & Jesse, M. (2017). Creating immersive and aesthetic auditory spaces in virtual reality. In *2017 IEEE 3rd VR workshop on sonic interactions for virtual environments (SIVE)*, Los Angeles, CA (pp. 1–6). IEEE.
Summers, C., Lympouridis, V., & Erkut, C. (2015). Sonic interaction design for virtual and augmented reality environments. In *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)*, Arles, France (pp. 1–6). IEEE.
Timmermans, H. (2015). Sound spatialisation from a composer's perspective. In *2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE)*, Arles, France (pp. 1–5). IEEE.
Välimäki, V., Parker, J., Savioja, L., Smith, J. O., & Abel, J. (2016). More than 50 years of artificial reverberation. In *Audio engineering society conference: 60th international conference: dreams (dereverberation and reverberation of audio, music, and speech)*, Leuven, Belgium. Audio Engineering Society.
van den Doel, K. (2005). Physically based models for liquid sounds. *ACM Transactions on Applied Perception*, *2*, 534–546.
van den Doel, K., Kry, P. G., & Pai, D. K. (2001). Foleyautomatic: Physically-based sound effects for interactive simulation and animation. In *Proceedings of the ACM SIGGRAPH 2001* (pp. 537–544). Los Angeles, CA.
van den Doel, K., & Pai, D. K. (1998). The sounds of physical shapes. *Presence: Teleoperators and Virtual Environment*, *7*(4), 382–395.
van Walstijn, M., & Mehes, S. (2017). An explorative string-bridge-plate model with tunable parameters. In *Proceedings of the international conference on digital audio effects (DAFx17)* (pp. 291–298). Edinburgh.
Xie, B. (2013). *Head-related transfer function and virtual auditory display*. Plantation, FL: J. Ross Publishing.
Zheng, C., & James, D. L. (2009). Harmonic fluids. *ACM Trans. Graphics*, *28*(3), 37.
Zheng, C., & James, D. L. (2011). Toward high-quality modal contact sound. *ACM Trans. Graphics*, *30*(4), 38.