ORIGINAL ARTICLE

FreesoundVR: soundscape composition in virtual reality using online sound repositories

Luca Turchet[1] · **Marco Carraro**[1] · **Matteo Tomasetti**[1]

Abstract

The intersection between sound and music computing and Virtual Reality (VR) has grown significantly over the past decades, amounting to an established area of research today. However, still scarce research has been conducted on the development of specific tools for sound design and composition. In this paper, we investigate a new way of exploring online sound repositories to retrieve sounds to be used in soundscape composition, which leverages the VR medium. Specifically, we created a VR system that allows users to search, download, and explore Freesound content in an immersive manner, as well as to use it for soundscape composition practices via a virtual digital audio workstation (DAW). The tags associated to a sound in the repository were converted into virtual objects and environments, which the user could navigate while listening to the sound. We conducted a user study with 16 composers where the developed system was compared against a conventional counterpart comprising the Freesound web version and the Audacity DAW. Overall, quantitative and qualitative results did not indicate a clear and generalized preference for a system over the other. The usability of the two systems along with their offered creativity support, cognitive workload and emotional impact were deemed to be at a comparable level. Nevertheless, the full potential of VR in creating novel compositional experiences also clearly emerged. Our study shows that VR is an effective medium to support users' creativity during the process of exploring and selecting sounds from an online repository as well as for composing a soundscape.

Keywords Musical XR · Composition · Soundscape · Internet of audio things

1 Introduction

Today various repositories of sonic content are available online (e.g., Soundly, Resonic, Splice), and some of them are freely accessible through Creative Commons licenses (Font et al. 2016). These repositories have found application in several contexts, such as sound design, film soundtracks creation, musical performance, musical composition, and auditory scene analysis research. One of the main sound-based practices in which sound repositories are utilized is soundscape composition (Truax 2008), where sound designers or composers create a soundscape by concatenating and/or superimposing different sounds, using a sound editing software. The term soundscape refers to all the sounds that can be heard in a specific location. This sonic environment is the aural counterpart of the term landscape referred to visually related items in an environment.

The conventional way of exploring an online sound repository consists of textual searches conducted on a web browser, where matches between the user queries and the tags indexing the repository return a list of sound file names (Font et al. 2013). In some cases, the returned file names are complemented with images of the waveform and spectrum of the sound. Virtual Reality (VR) can provide new ways to interact with online sound repositories and support soundscape composition. Indeed, nowadays VR is increasingly used in applications supporting creativity both in academic and industrial settings. In particular, this medium has made inroads into musical practices such as composition and performance (Turchet et al. 2021). However, to the best

✉ Luca Turchet
  luca.turchet@unitn.it

  Marco Carraro
  marco.carraro@unitn.it

  Matteo Tomasetti
  matteo.tomasetti@unitn.it

1 Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, 38123 Trento, Italy

of authors' knowledge, the exploration of an online sound repository using Virtual Environments (VEs) has been largely overlooked. Along the same lines, the practice of soundscape composition in VR leveraging online sound repositories has been scarcely addressed. In general, still few VR-based tools are available to composers and sound designers to support their creative processes.

In this paper, we aim to bridge these gaps by proposing a new way of exploring online sound repositories and performing soundscape composition that leverages the VR medium. Our investigations were driven by two main research questions:

1. What are the added value and limitations of exploring the content of a sound repository in VR compared to the conventional use of a browser?
2. To which extent VR is an effective medium to support users' creativity during the process of composing a soundscape with content retrieved from a sound repository?

To address these questions, we created a VR system that allows users to search, download, and explore Freesound content in an immersive manner, as well as to use it for soundscape composition practices. The tags associated to a sound in the repository were converted into virtual objects and environments, which the user could navigate while listening to the sound. Our hypothesis was that through an immersive experience the user would explore the repository and perceive the retrieved sonic content in more engaging ways compared to the conventional browser-based interaction. As such, this novel exploration and understanding of the sounds would lead to new ways of composing soundscapes and intend the practice of soundscape composition.

This work falls in the remits of the Internet of Audio Things paradigm, which deals with sounds in networked contexts (Turchet et al. 2020), and extends this field by leveraging the new possibilities offered by the VR medium.

2 Related work

2.1 Soundscape composition

The term "soundscape" refers to all the sounds that can be heard in a specific location. This sonic environment is the aural counterpart of the term landscape referred to visually related items in an environment. Research on soundscapes has a long and established history. Studies on real soundscapes started with R. Murray Schafer, among others, in late sixties (Schafer 1977) and continued by focusing mostly on musical applications, with pioneering works of Barry Truax (Truax 1992, 1996). The term "Soundscape composition"

refers to a sound-based art form that concerns the creation of sonic environments (Westerkamp 2002; Drever 2002; Truax 2008). This art form has grown from acoustic ecology (Wrightson 2000) and soundscape studies (Westerkamp 2002).

Composed soundscapes are used widely in various contexts, including movies (d'Escriván 2009; Leonard and Strachan 2014), music performances (Truax 2008; Freeman et al. 2011), artistic installations (Chapman 2009; Koutsomichalis 2013), and VEs (Eckel 2001; Turner et al. 2003; Turchet and Serafin 2013). To date, soundscape composition is facilitated by the availability of freely available online sound repositories as well as high-quality commercially available sound effects libraries, conceived especially for creating environmental sounds in movies. Various authors have developed systems conceived for soundscape composition that leverage such repositories. These include real-time (Valle et al. 2009), interactive (Verron et al. 2009), non-interactive (Schirosa et al. 2010; Thorogood and Pasquier 2013), automatic (Valle et al. 2014), voice-based (Turchet and Zanetti 2020), and tangible systems (Huang et al. 2015).

2.2 Freesound and its artistic use

The Audio Commons Initiative is a recent endeavor aiming to bridge the gap between audio content producers, providers and consumers through a web-based ecosystem (Font et al. 2016). The approach combines techniques from Music Information Retrieval (to extract creative metadata to automatically annotate audio content) and the Semantic Web (to structure knowledge and enable intelligent searches). Sound repositories part of the Audio Commons ecosystem provide access to audio data through user-facing and application programming interfaces (APIs).

One of the most popular and freely available online repositories which is part of the Audio Commons ecosystem is Freesound[1], a collaborative repository of audio samples developed at and maintained by the Music Technology Group of Universitat Pompeu Fabra (Font et al. 2013; Fonseca et al. 2017). The Freesound database provides a collection of near 500k Creative Commons musical and nonmusical sounds contributed by a community of thousands of people around the world. In Freesound, the available metadata information about the sounds depends on what has been provided by authors during uploads including tags, descriptions or file names (Favory et al. 2018).

Freesound enables designers to create third-party applications exploiting its audio content in live contexts by granting access to the database through a REST API (Akkermans et al. 2011). Different authors have investigated the use of

---

[1] http://www.freesound.org..

such API for artistic purposes: Skach et al. (2018) proposed a system for body-centric sonic performance that allows performers to manipulate sounds retrieved from Freesound through gestural interactions captured by textile wearable sensors; Turchet and Barthet (2018) developed a system where a smart mandolin interacts with audience members' smartphones that generate musical accompaniments based on Freesound content; Font (2021) proposed an open-source hardware music sampler that uses content retrieved from Freesound; Turchet and Zanetti (2020) presented a voice-based system that allows to perform soundscape composition using a commercial voice-based interface.

2.3 Composition in VR

The intersection between sound and music computing and VR has grown significantly over the past twenty years, amounting to an established area of research today. Concerning composition, various authors have developed systems for this musical practice (Zappi et al. 2012; Buckley and Carlson 2019; Ciciliani 2020; Carey 2016; Costa et al. 2019; Graham and Cluett 2016; Moore et al. 2015). However, still scarce research has been conducted on the development of specific tools for composition and sound design. The analysis of the field recently conducted in (Turchet et al. 2021), revealed that to date only a limited number of VR systems directly target composers. Nevertheless, such analysis suggested that VR-based tools can be effective in supporting compositional processes as evidenced in (Buckley and Carlson 2019) and (Ciciliani 2020). The present study aims at answering the call of Turchet et al. (2021) to develop new software tools capable of supporting creativity in VR in order to progress the field.

3 The FreesoundVR system

3.1 Design

The interaction design process followed several cycles of design-prototyping-evaluation and was conducted not only by the three authors of this paper, but also by involving an experienced composer with expertise in VR as well as in Freesound. Such a composer was not involved in the main evaluation experiments reported in Sect. 4.

The main idea beyond our study was to translate tags of sounds retrieved from online repositories into virtual elements to be placed in VEs. This would have allowed users to explore the content of a sound file in an unprecedented way, that is in an immersive setting. Such a novel exploration would have been beneficial for generating compositional ideas while composing a soundscape. Therefore, we designed a VR system that enables users to search,

explore and download content of Freesound, as well as to use the downloaded sounds for composing a soundscape. Each downloaded sound gets transformed into an immersive VE that the user can navigate. Such a transformation is performed automatically on the basis of the tags associated to the sound and through a mapping between such tags and virtual elements. We called the resulting system "FreesoundVR", as a VR counterpart of the Freesound service[2]

The Freesound database is very large, encompassing thousands of sounds and hundreds of tags. As the developed prototype was mainly intended to be a proof-of-concept, for feasibility reasons we made some design choices. Firstly, we focused only on 9 macro categories of sounds and on 3 sounds for each category, for a total of 27 sounds. Specifically, the 9 macro categories were: Game, Horror, Sci-Fi, Cars, City, Dance, Animals, Nature, and Birds. They were selected because they are among the sounds most available and most searched on Freesound. Secondly, we decided to preselect the possible 27 sounds, by retrieving their unique Freesound ID. For each of these sounds, we then limited the number of tags selecting the most common ones, which led to a total of 44 tags (and therefore, to as many virtual objects). For each sound, we created a VE encompassing the virtual objects associated to the tags, leading to 27 VEs.

Concerning the soundscape composition process following the selection of the sound files, we adopted the Digital Audio Workstation (DAW) paradigm that is common in the soundscape composition practice. Notably, our aim was not to develop an advanced DAW in VR, since our goal was to provide users with the most common functionalities allowing them to create a basic soundscape. We identified such functionalities in the possibility of moving the sounds along a timeline in each track, controlling the volume of each sound, and providing basic playback tools for each sound as well as mute/solo options.

3.2 Implementation

FreesoundVR was realized as an app for the Oculus Quest 2. It was developed in Unity 3D and made use of a Python client that leveraged the Freesound APIs (Akkermans et al. 2011). Such a client[3] was necessary to perform the interaction with the Freesound repository, as a Freesound API is not available in the native code of Unity 3D, i.e., C#. For this purpose, the Python for Unity package was utilized. The Python client allowed to establish an indirect communication between the VE and the Freesound server, in order to perform real-time downloads of the selected sound files

---

[2] A video of the system in action is available at https://youtu.be/q30iAq1YU8U.
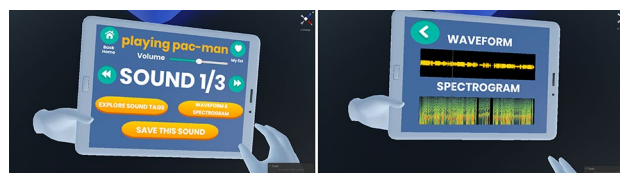
[3] The client is available at: https://github.com/.

Fig. 1 An example of a VE generated from a sound belonging to the Animals category

Fig. 2 The main room

Fig. 3 The tablet used for sound selection and information display

and obtain the main information, such as the title and the descriptive tags of the sound itself. The retrieved tags were then automatically transformed into 3D models of virtual objects (see Fig. 1), using Unity 3D Prefabs. The 3D models for these Prefabs were obtained from Sketchfab, a platform for publishing, sharing, exploring, buying and selling VR and AR content. For the acquired models, the necessary step at a later stage was to resize and place them within each created VE representing a sound. In all the created VEs, the position and temporal evolution of the virtual objects were set to generate landscapes that are coherent with the real-world counterpart.

The FreesoundVR system was designed to be controlled via handheld controllers. It was composed by three rooms, the Main Room, Mixing Room and New Personal Sound Room, plus the 27 VEs displayed as a result of the conversion of the sound tags. The purpose of the Main Room is to allow users to select the sounds to be retained for sound-scape composition (see Fig. 2). In the Main Room, the user could select the various macro categories of sounds, each represented by the metaphor of a vinyl disc to be taken from a wall and put on a vinyl record player. Upon this selection, a VE corresponding to one of the three sounds preselected for such category was displayed at visual level, while the sound

was also played. At this stage, the user could decide to keep the sound or pass to the next VE representing another sound of the category. These interactions were managed through a virtual tablet (see Fig. 3), from which the user could also display the sound tags as well as the related waveform and spectrogram (as a parallel with the information available on the original browser-based Freesound version).

The sounds that the user can decide to retain are a maximum of 6, and are represented inside one of 6 semitransparent spheres suspended in mid-air along a wall of the Main Room through a small thumbnail (see Fig. 4). Once the sound selection has been made, the user can interact with the spheres to play the selected sounds and, therefore, listen once more to them before entering in the Mixing Room. Specifically, the sound reproduction starts when the user inserts a virtual hand inside a sphere, and stops when the hand is removed. Both hands can be used for this purpose, leading to two simultaneously played sounds. The list of saved sounds, moreover, is always visible to the user through the tablet, who can also decide to remove a saved sound.

The Mixing Room (see Fig. 5) contains a 3D DAW. Each selected sound is displayed on a track with a corresponding waveform (a maximum of 6 tracks is available). Each sound can be moved along the track timeline by a "drag and drop" movement, played individually or jointly, muted and altered in volume. The drag and drop actions necessarily require physical movements of the user, and to decrease the consequent effort, the user is equipped with a remote control that allows him/her to move the entire DAW. The remote control is also equipped with buttons related to the main functions of the DAW such as the Play, Pause, and Stop, as well as the command to save the DAW configuration. Finally, the user can perform a mixdown of the tracks, obtaining a new .wav file. This process could take up to 3 minutes. During this period the user is informed about the remaining time needed to complete the operation, and is provided with some text to read that contains information about the Freesound platform.

The mixdown file results in a new VE, the New Personal Sound Room (see Fig. 6). This room is the result of the combination of all the virtual objects associated to the sounds selected. Therefore, it will change based on the individual sounds selected during a soundscape composition session. Inside this room the user, through a tablet, can listen to the created sound and control its volume and reproduction.
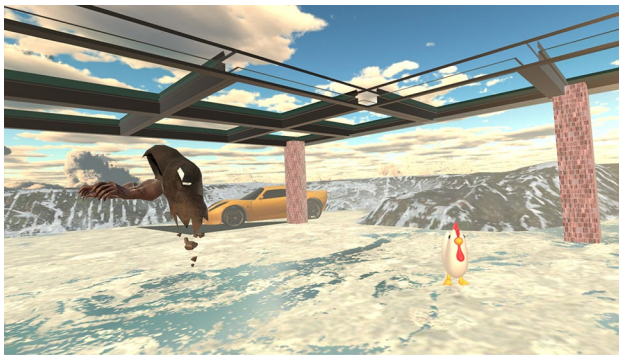
Fig. 4 The sound spheres part of the Main Room
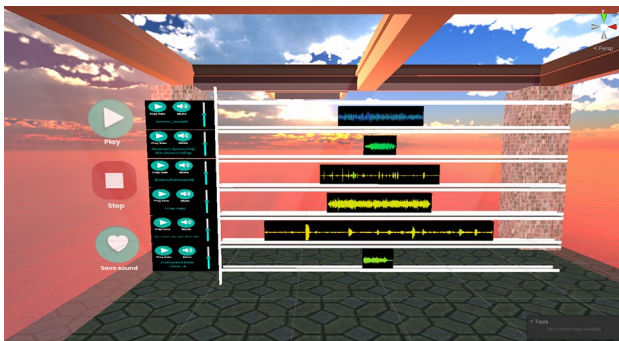
Fig. 5 The mixing room

Fig. 6 The new personal sound room, which is the result of the mixdown of sounds from the category Animals, Horror, Cars and City

Moreover, the user can view all the tags resulting from the individual sounds that made up the new composition, the name of the rooms from which they were saved, and the titles of the sounds.

It is worth noticing that the user can always access the various sound-related VEs during the actual composition phase, and not just before. This is ensured by the possibility to exit the Mixing Room and entering the de-sired VE. This is an action that can be performed easily and quickly. However, we opted to avoid showing during the composition

phase the visual elements of the various VEs related to the selected sounds, i.e., directly in the Mixing Room. This choice was due to our precise goal to avoid increasing too much the cognitive load of the user during a complex activity such as that of composition. Notably, according to the established workflow, the user at the time of sound composition will come from a period of sound research in which s/he was immersed in the various VEs. Moreover, in the Home, the transparent spheres representing each selected sound served as a further anchor to memorize/remember not only the sounds but also the related VEs. Furthermore, through the hand-based interaction with the spheres, the user is allowed to preliminary compose the soundscape before entering the Mixing Room, as the system enable to listen up to two sounds simultaneously. Finally, by interacting with the New Personal Sound Room after the mixdown, users can keep exploring the sound content, which this time will be the result of the superimposition of sounds and VEs. Users have the ability to appropriate this process in the most convenient way for them, such as creating the soundscape iteratively by mixing progressively an increasing number of files and at each stage exploring the resulting sonic space.

3.2.1 Details on the VEs generation process

VEs were generated based on the descriptive tags of each sound. Within Freesound, each sound is characterized by a variable number of descriptive tags, simple words that summarize the content of the sound file. Some of these tags may represent abstract definitions, while others are references to concrete objects. The generation of VEs was based on the latter type of tags. Within our program, we relied on a small dictionary, which mapped a descriptive tag to an object that can be instantiated within a scene (stored in the form of a 3D object Prefab). Whenever the user enters inside a scene and a sound is played, the program receives from Freesound the corresponding set of descriptive tags and matches them against the stored dictionary. Each match will result in the instantiation of a Prefab within the scene. It follows that the more a sound will be described on Freesound through these tags, the more accurately it will be representable within a VE.

Therefore, a VE can be composed of a variable number of Prefabs. To ensure that they do not conflict with each other (for example by overlapping) they were manually inserted within a "tester" scene in an initial design phase. Once the animations for each element were created, and suitable placements were studied, these characteristics were automatically saved within the information of the Prefab itself. Consequently, whenever a Prefab is instantiated within a VE, it assumes the placement characteristics with which it was stored. In essence, all elements are placed automatically after passing a necessary first manual placement phase.

In our design, the generated VE is closely related to the sound file the user is listening to at that moment. In the context of the experiment, each of the nine sound categories selected has three sound files. We selected such files because they had extremely personalized descriptive tags that were different from each other. This ensured that each file produced a different environment even within the same sound category.

For the New Personal Sound Room, the matter is slightly different. Here there is the need to combine elements of radically different sounds, not only acoustically but also in terms of visual representation. For this reason, in the New Personal Sound Room, descriptive tags were no longer used to instantiate elements in the scene. The program connects to each saved sound a single element present in its VE, i.e., the most characteristic element, the same one represented within the sound spheres in the Home. Such an element was automatically selected based on the dictionary, which prioritized some elements rather than others. This made it possible to have a single element for each sound to be represented within the new scene created, with significantly greater ease of management. Therefore, the number of elements present within the New Personal Sound Room is related to the number of sounds with which the user has decided to work on his composition. Two selected sounds will produce two elements, while using all available slots will produce six elements.

4 Evaluation

4.1 Participants

A total of 16 participants took part to the evaluation (15 males, 1 female, aged between 23 and 45, mean age = 29.6). All participants were musicians with expertise in music technology, had previous experience in using VR applications and in soundscape composition practices, as well as had previously used Freesound and Audacity (a free and widely used DAW). Participants were from different nationalities and were distributed worldwide.

4.2 Procedure

The evaluation procedure was devised to answer to the two research questions listed in Sect. 1. The experiment was divided into 2 parts, each testing a different system to download sounds from the Freesound online repository and then to mix the downloaded sounds to create a soundscape.:

FA ("Freesound + Audacity"): this system comprises the Freesound version in the browser from which users can download sounds, and the Audacity DAW, which allows the user to mix the sounds;

FVR ("FreesoundVR"): this system is the VR application described in Sect. 3.

Participants were tasked to compose, in both systems, one soundscape, using 6 sounds belonging to one of the 9 categories (Game, Horror, Sci-Fi, Cars, City, Dance, Animals, Nature, Bird). At the beginning of the experiment, participants were debriefed about the procedure and they underwent a familiarization phase where they tried each system before accomplishing the task. After the completion of the task for each system, participants were asked to fill a questionnaire devised to assess the usability of the system, investigate the degree of creativity fostered by the system, assess the resulting emotional impact, and understand the perceived cognitive workload. Specifically, the questionnaire comprised: i) a set of ad hoc open-ended questions about the experience in interacting with the system; ii) the raw NASA TLX (Hart 2006) to measure overall workload and the six individual factors which include mental demand, physical demand, temporal demand, performance, effort, and frustration; iii) the Self-Assessment Manikin (SAM) (Bradley and Lang 1994) to assess the experienced arousal, valence and dominance; iv) the questionnaire to calculate the creativity support index (Cherry and Latulipe 2014); and the System Usability Scale questionnaire (Brooke 1996). Finally participants were given the possibility to leave an open comment.

Half of the participants started with the system FA and then used system FVR, vice versa the other half. In both systems, participants were let free to select the sounds, no specific order was provided. Seven participants made the experiment in a laboratory of University of Trento, while nine at their home. All participants used an Oculus Quest 2. Before using the FVR system participants were asked to watch a video tutorial illustrating the functionalities of the system. Participants took on average one hour and a half to complete the experiment. The procedure, approved by the local ethics committee, was in accordance with the ethical standards of the 1964 Declaration of Helsinki.
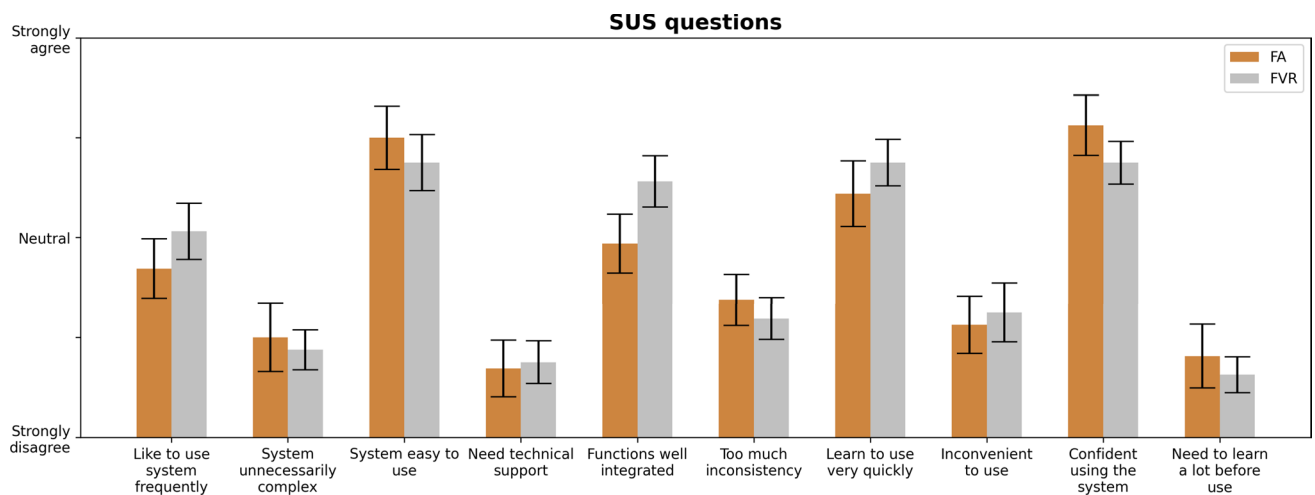
4.3 Quantitative results

4.3.1 Raw NASA TLX

The total workload index for system FA was 245.62 (SD = 64.12), for system FVR was 274.37 (SD = 19.22). A Wilcoxon–Mann–Whitney test showed that
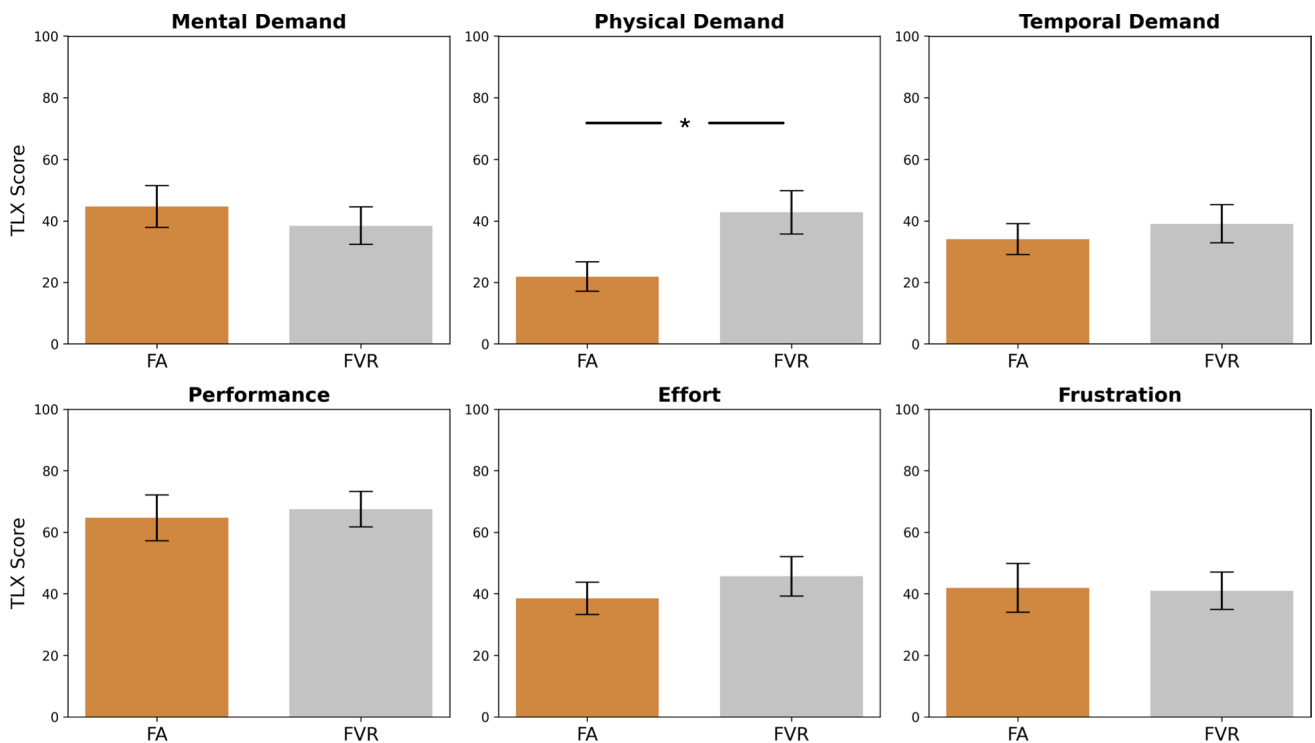
Fig. 7 Mean and standard error for each individual subscale of the NASA TLX

Fig. 8 Mean and standard error of the system usability scale topics for both systems

there is no significant difference between the systems. Figure 7 shows the mean and standard error for each individual subscale of the NASA TLX. Using a Wilcoxon–Mann–Whitney test, the subscale physical demand for FA was rated as significantly lower than that for FVR (W = 66.5, $p < 0.05$).

4.3.2 System usability scale

The SUS metric assesses the usability of a system on a scale from 0 to 100. As a point of comparison, an average SUS score of about 68 was obtained from over 500 studies. System FA obtained a mean SUS score of 67.96 (95% confidence interval: [55.30; 80.63]) which is around average. System FVR obtained a mean SUS score of 70.46 (95% confidence interval: [62.48; 78.45]) which is above average. An

Table 1 Average CSI results for the Freesound + Audacity (SD reported in brackets)

Creativity factor	Avg. factor counts		Avg. factor score		Avg. weighted factor score	
	FA	FVR	FA	FVR	FA	FVR
Enjoyment	1.81 (0.8)	2.75 (1.52)	10 (6.23)	**12.25** (4.57)	17.25 (14.1)	29 (15.41)
Exploration	**4.37** (0.59)	3.18 (1.23)	11.81 (5.47)	8.87 (4.72)	**52.5** (24.69)	31.06 (23.06)
Expressiveness	3.31 (1.26)	**3.43** (1.22)	10.87 (6.23)	8 (5.18)	34.5 (24.13)	30.06 (24.23)
Immersion	2.37 (0.78)	2.75 (0.9)	9.56 (4.75)	10.31 (6.16)	22.18 (12.24)	27.75 (20.2)
Results Worth Effort	3.12 (0.99)	2.87 (1.36)	**12.68** (5.54)	10.68 (5.57)	41.62 (25.22)	**31.62** (24.44)

The highest average value is reported in bold in each column. The mean CSI score for FA is 56.02 (SD = 26.41), for FVR is 49.83 (SD = 23.11). Ranges: Avg. factor counts (0 to 5), avg. factor score (0 to 20), avg. weighted factor score (0 to 100)
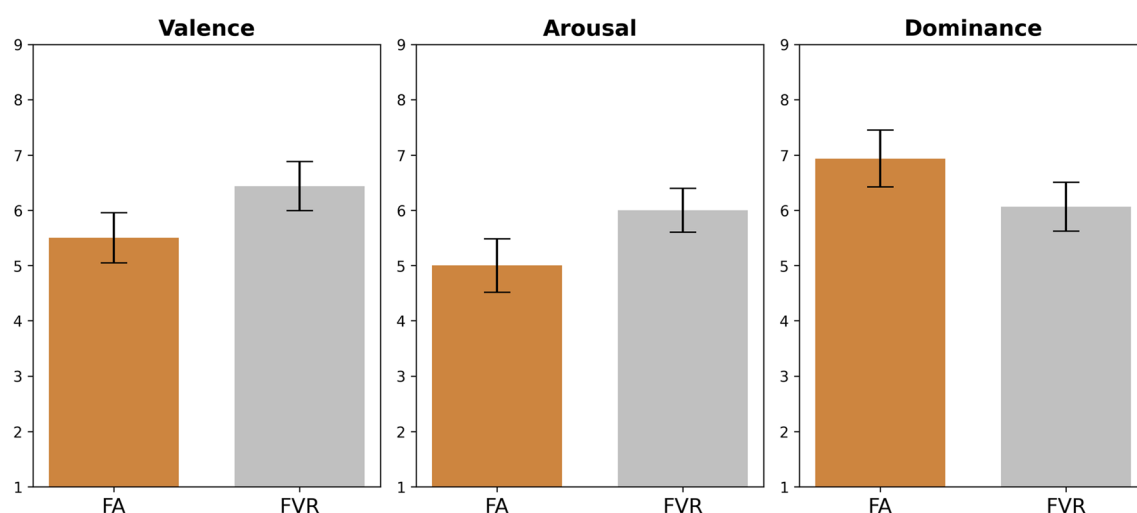
Fig. 9 Mean and standard error of the SAM questionnaire for both systems

analysis conducted with the Wilcoxon–Mann–Whitney test revealed that such differences between the systems were not statistically significant. Figure 8 shows the breakdown of the result across the SUS topics. A Wilcoxon–Mann–Whitney test showed no statistically significant difference between the systems for any of the SUS topics.

4.3.3 Creativity support index

The CSI metric, ranging in [0,100], enables to assess the ability of a tool to support the open-ended creation of new artifacts. However, the CSI involves a scale related to collaboration, which is an aspect not present in the systems under study. Therefore, such scale was defaulted to 0 as indicated by the authors of CSI (Cherry and Latulipe 2014). The FA system obtained a mean CSI of 56.02 (SD = 26.41), the FVR system obtained a mean CSI of 49.83 (SD = 23.11). An analysis conducted with the Wilcoxon-Mann-Whitney test revealed that such differences between the systems were not statistically significant. Table 1

presents the average CSI results broken down into factor counts (the number of times a creativity factor was judged more important than another for the task, as based on paired comparisons), factor scores (the ratings of the various factors irrespective of their importance for the task), and the weighted factor scores, which combine the factor counts and scores to make it more sensitive to the factors that are the most important to the given task.

4.3.4 Self-assessment manikin

Results of SAM for both systems are illustrated in Fig. 9. An analysis conducted with the Wilcoxon–Mann–Whitney test revealed that any differences between the systems were not statistically significant, for valence, arousal, and dominance.

4.4 Qualitative results

Participants' answers to the open-ended questions were analyzed using an inductive thematic analysis (Braun and Clarke 2006). Such analysis was conducted by generating codes, which were further organized into the following themes that reflected patterns.

4.4.1 Themes regarding FA

Easiness. Nine participants commented that the process of composing soundscapes via the FA system was straightforward. They reported that it was easy to search and download the wanted sounds and build an audio track (e.g., *"I found myself at ease. The ability to decide the length of the sounds to download, the ability to download high-quality uncompressed files and the huge choice of sounds won me over. It allowed full involvement and full artistic expression, obtaining the exact result I had set for myself given the task constraints."*).

Freesound expressivity. Seven participants appreciated the fact that Freesound contains a large variety of sounds to choose from, which allows them to support their creativity during the process of soundscape composition (e.g., *"The experience was quite good, the Freesound's library of sounds is so huge and I spent so much time to choose the right sounds for my composition. I was satisfied of the result in the end."; "There is a lot of variety on Freesounds, so using their platform allowed me to be creative with choosing sounds."*). These comments, however, are likely to be originated from the fact that the involved participants were already acquainted with the Freesound website, although in the present study only a restricted number of sounds were involved. Nevertheless, such comments also provide an indication of the expressive potentialities of mapping the entire Freesound catalog into VEs.

Limits of Audacity. Four participants commented that Audacity was not the best tool to perform soundscape composition, given some limits in the features it offers (e.g., *"Using Audacity was cumbersome and frustrating. I'm used to newer DAWs with more initiative UI and controls."; "For those who are trained on Audio Editing, Mixing and Composing, Audacity could be a too much basic DAW, but potentially it has all the tools to make a good composition."*).

Integration needs. Five participants commented that the FA system lacked a proper integration between Freesound and Audacity, which would have facilitated the soundscape composition process by reducing workflow interruptions (e.g., *"It would be great to avoid downloading the sounds and then loading them in Audacity: I would love to listen to a sound on Freesound and then directly drag and drop it into Audacity. An integrated system would support this process."*). Some suggested to have a web-based DAW that is closely integrated with Freesound, so that the user can find sounds and compose the output in a single page.

4.4.2 Themes regarding FVR

Enjoyment. Seven participants commented that overall their experience was positive, interesting, enjoyable and fun (e.g., *"I loved the visuals of the soundscapes when retrieving samples. They were super cool to look at and made the experience enjoyable."; "I enjoyed seeing a new way of creating sound in VR."*). Three participants commented that the system was intuitive to use (e.g., *"I watched the tutorial video only one time, and once inside FreesoundVR, it was quite intuitive to use."*). Three participants reported to have appreciated the concept of Freesound VR.

Limits of the virtual DAW. Ten participants reported that the affordances of the virtual DAW were too limited to support properly their creativity (e.g., *"The main limitation is that the DAW is currently not expressive enough to support the interactions I would need."*). Firstly, they lacked more sound processing controls (e.g., *"I would have loved some effects like reverb, delay, granulator, anything that would give me the tools to put my own flavor on the samples."; "I would like to have more functions in the DAW, such as duplicating clips, cropping them and fading clip beginning/ends."*). Secondly, they lacked the possibility to spatialize sounds in 3D (e.g., *"It would have been fun to be able to spatially mix since it is a visual soundscape experience."; "It would be cool to mix spatially by actually moving the sounds around 3D space in the mix room."; "I like a lot the idea behind FreesoundVR but I would like that I can move inside the world and the sound would be spatial."*). Thirdly, they complained about a lack of information about the sounds, which is typically available in conventional DAWs (e.g., *"During the selection of the track I couldn't visualize important informations like how long is a particular track."*).

Limits in virtual navigation. Two participants who performed the test at their own home reported to have experienced issues in navigating the VEs. These issues were due to a limit in the area available in the real world, which was smaller than that needed for performing the interactions that were designed (e.g., *"I could not access easily the mixing room because I have not enough space in my room to walk that far."; "Having neither a very long link cable nor a very large environment in which to move, I missed the possibility of having a movement system to interact more effectively."*).

4.4.3 Themes regarding the comparison between systems

Benefits of immersion. Eleven participants commented that FVR provides various benefits compared to FA. Firstly, according to six participants the immersive nature of the

experience was more fun, enjoyable, and allowed them to be more engaged in the sound search and composition processes (e.g., *"In FreesoundVR the involvement was full and much more emotional.";* *"I loved the immersiveness provided by the 3D worlds, this is something which can make the sounds be seen under a different perspective. Sounds felt more "alive" to me.";* *"Searching for sounds was way more engaging that on the browser, given all the different environments in which one is immersed as one searches.").* Secondly, for five participants the system allowed a different perception of the sounds and as a consequence of their selection (e.g., *"The impact that the image had in making me perceive sounds was very important. I felt immersed and it was pleasant and curious. The interaction made me feel part of the world.";* *"The visual images would likely influence my choice of sounds in VR, i.e., I might be more inclined to choose a sound if I like the visual.").* Thirdly, FVR was perceived by four participants to have a better creativity support (e.g., *"I think that working in VR could make you more creative than working in a traditional way.").* Fourthly, two participants commented that FVR was more usable than FA (e.g., *"the 3D interface is much more usable compared of the same amount of tracks displayed on a flat screen.").*

Integrated functionalities. Six participants commented that the main added value of FVR over FA was the integration of the system for sound retrieval and mixing (e.g., *"I think the ease of use it a great part of it. Having everything in one place can be really helpful for beginners.").*

Drawbacks of the virtual DAW. The virtual DAW was deemed too simple by six participants, who felt limited in the affordances provided compared to those of Audacity. The causes have been listed in the previous section. Such limits were reported to have impacted negatively their overall experience of the system. However, they also reported that improving such an aspect would lead them to prefer FVR over FA (e.g., *"FreesoundVR doesn't have the effect library that Audacity has. If FreesoundVR had more effects and mixing functionality, there would be no question between the two. FreesoundVR is more fun and creative.";* *"It would not be that hard to bring FreesoundVR in par with the full experience of Freesound + Audacity on PC and this would result in a very interesting tool.").*

Limits of number of available sounds. Four participants felt limited in the possibility to explore and choose the sounds. Indeed, the pool of 27 sounds provided was deemed not sufficient and they would have loved to have the whole Freesound dataset available (e.g., *"Sound choice is limited, I would like that there was more sounds.";* *"The Freesound browser has more samples to choose from. If FreesoundVR had a much bigger sample library, I would definitely use it more often to create.").*

Rendering time. Two participants reported to have felt rather annoyed by the 3-minute rendering time caused by the generation of the mixdown (e.g., *"The main issue is the Offline Render which obligate user to wait").*

Issues with HMD. Three participants commented that the use of an HMD-based system such as FVR for the activity of soundscape composition could not be ideal for long sessions, which usually happen in such a practice (e.g., *"The use of the headset is still too difficult and uncomfortable").*

5 Discussion and conclusions

In this paper, we investigated a new way of exploring online sound repositories for soundscape composition practices by leveraging the VR medium. The use of VR represents an element of novelty in this space, as previous systems developed for the same purpose utilized different approaches, such as desktop computers (Schirosa et al. 2010; Thorogood and Pasquier 2013; Valle et al. 2014), voice-based interfaces (Turchet and Zanetti 2020), or tangible objects (Huang et al. 2015).

The conducted user study focused on a comparison between the developed system and the conventional approach that involves the Freesound web service and a DAW. Overall, the quantitative and qualitative results did not indicate a clear and generalized preference for a system over the other. The usability of the two systems along with their offered creativity support, cognitive workload and emotional impact were deemed to be at a comparable level.

The user study revealed that participants learned to use the FVR system easily and quickly given a short training period and that they were able to achieve good results. FVR was not deemed less intuitive and easy to use than FA. As shown in Fig. 7, participants performed their tasks with a similar accuracy, level of frustration, time and degree of effort. The only difference concerned the physical demand, which was higher for FVR. This is a plausible result as the use of an HMD and involving gestures via the VR controllers takes more physical movements than those occurring when sitting on a chair and operating with mouse and keyboard in front of a screen.

The three SAM dimensions (valence, arousal and dominance) did not differ between the two systems. Nevertheless, from Fig. 9 it is possible to notice some tendencies. Valence and arousal received on average higher scores in FVR compared to FA, which is an indication of the level of enjoyment and engagement experienced by participants. It is plausible to hypothesize that such differences did not turn to be statistically different because of the prototype limitations reported by the participants (e.g., limits of the DAW, issues with navigation, restricted sound availability, rendering time, and lack of comfort in the prolonged use of the HMD). These were reported to have negatively affected their experience and limited the artistic quality of the generated

output. Nevertheless, participants also agreed that if such issues could be solved they would likely prefer to use FVR compared to FA.

The creativity support offered by FVR did not significantly differ from that of FA. This is also supported by the comments of some participants who deemed that essentially nothing changed in the compositional process. The composers were capable of achieving their own musical goals in both systems. However, one of the main issues reported by participants for FVR was the lack of the wide range of sounds to choose from, which is instead available in FA. This was deemed to limit their creativity. On the other hand, interestingly a few participants explicitly mentioned that FVR supported and stimulated better their creativity compared to the use of FA.

In general, all participants seemed to have genuinely appreciated the concept underlying FVR, where a sound could be explored not just via tags and in 2D, but via a full 3D visual immersion. This was considered a useful added value compared to the traditional approach, leading to a fun, interesting, enjoyable and novel experience. Some participants referred to the sounds as "more alive" in the created virtual worlds, and that the immersive nature of the FVR system led them to perceive the sounds in a different, unprecedented way. This is one of the novel aspects of our work, which highlights affordances, interactions, and perceptions only possible with VR rather than those achievable with more conventional media.

Taken together, the results reported in this study show that VR is an effective medium to support users' creativity during the process of composing a soundscape with content retrieved from a sound repository. This is in line with other studies available in the literature, which highlighted how VR-based tools can be effective in supporting compositional processes (see e.g., (Buckley and Carlson 2019; Zappi et al. 2012; Ciciliani 2020). The advantages offered by VR over traditional screens and digital media lie in its ability to create immersive and magical experiences. Our study suggests that this holds true also for the process of exploring a sound repository and using creatively the retrieved sounds. What is interesting from the comments of some participants is that VR has the ability to change the perception of a sound, and its immersive nature can provide different benefits to compositional activities.

Interestingly, for some participants, the effect of visual representation seems to be orienting or influencing the choices of a sound against another. Nevertheless, this may be due to personal interpretations of the hedonic qualities of a VE corresponding to a sound rather than more objective features of the VE itself. Nonetheless, this observation pinpoints to the importance of creating appropriate VEs that represent virtually a given sound and enable its exploration in novel ways. To our best knowledge, this is an aspect that at present has not been investigated by research, and that in this study has been addressed by the creativity of the designers. Future research could focus on the definition of some guidelines capable of supporting designers in the tackled process of mapping sound tag to virtual elements in VEs.

The approach adopted in the system presented here is based on the mapping of sound descriptors to virtual elements. Such an approach is different from that of other VR systems for composition previously developed. For instance, the attention's researcher has focused on porting in 3D the DAW paradigm, creating novel 3D visualizations of sound-related information, aiming at exploiting tridimensionality to achieve more efficient interactions with the visualized sonic content (see e.g., (Polfreman 2009; Barri 2009)). Conversely, FreesoundVR supports creativity by immersing the composer in a VE that represents the sound at hand, turning the sound in an "alive" experience. In a different vein, other authors focused on immersing the composer in a VR representation of the scores, while listening to the corresponding music (Masu et al. 2020). The aim was to facilitate users to develop a personal relationship with both the system and the score. This goes in the same direction of FreesoundVR where creativity is supported by building a new relationship between the sonic content and the user, by providing an immersive experience. Nevertheless, the main aim of FreesoundVR is that of fostering creativity by immersing users during sound exploration. Another example is the system VrGrains developed by Zappi et al. (2012) uses concatenative synthesis, where audio segments are combined on the basis of descriptors. The descriptors are spatially organized, based on their value, in a 3D environment that the user can navigate to explore the sonic space. Such descriptors, however, were related to the actual features of the sound (e.g., loudness, pitch, periodicity) rather than on the high-level semantic of the content (e.g., birds, forest, wind) like in FreesoundVR. Moreover, FreesoundVR relates on the superimposition of sound files like in the conventional soundscape composition, rather than on the use of concatenative synthesis. Nevertheless, common to VrGrains is the adoption of natural interactions to support creativity in VR, in place of the mouse-screen paradigm that characterizes other conventional composition systems.

It is worth noticing that our study presents some limitations. First, the study involved a restricted number of sounds. This was due to the fact that for each of them one needs to utilize a 3D asset that fits. This can result in a long and challenging development process. Our aim with this study was simply to achieve a proof-of-concept prototype; therefore, we limited the sounds to a number that we deemed sufficient for demonstrating the functioning and the potential expressivity of the system. Nevertheless, the system's structure is designed to be able to handle a much larger number of sounds. Another limitation of our study is represented by

the fact that in an ideal case a unique program combining Audacity and Freesound would have represented a better baseline to compare against FreesoundVR. However, such a program to our best knowledge does not exist. Moreover, we aimed at using the tools that composers and sound designers typically utilize in their workflows for soundscape composition.

Several avenues are possible for future work to extend the development of the FVR system and improve the quality of the interactions available to the users. These are the direct result of the features requests emerged from the conducted user study. Firstly, we plan to improve the virtual DAW, extending it with the controls suggested by participants (especially effects, advanced editing tools), as well as with information about the sounds with a higher level of detail (e.g., duration of the sound). Secondly, we plan to integrate in FVR a system for the 3D sound spatialization. This would improve the temporal and spatial coherence of the virtual objects with respect to the heard sounds. Thirdly, we plan to create a version of FVR that can be navigated within the boundaries of a restricted physical world, in order to make FVR more accessible. Finally, we plan to extend the number of sound tags and categories to be automatically rendered into virtual objects and environments, so to provide FVR with a range of possible choices similar to that offered by Freesound.

Declarations

Conflict of interest The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The submitted article was not supported by any funding agency.

Informed consent We administered an informed consent to all participants involved in the reported study

Human and animal rights The research involved human participants

References

Akkermans V, Font F, Funollet J, de Jong B, Roma G, Togias S, Serra X (2011) Freesound 2: an improved platform for sharing audio clips. Proceedings of the international society for music information retrieval conference

Barri T (2009) Frameworks 3D: composition in the third dimension. Proceedings of the international conference on new interfaces for musical expression, pp. 264–265

Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exper Psych 25(1):49–59

Braun V, Clarke V (2006) Using thematic analysis in psychology. Qualitat Res Psychol 3(2):77–101

Brooke J et al (1996) Sus-a quick and dirty usability scale. Usabil Evaluat Ind 189(194):4–7

Buckley Z, Carlson K (2019) Towards a framework for composition design for music-led virtual reality experiences. IEEE conference on virtual reality and 3d user interfaces pp. 1497–1499

Carey B (2016) Spectrascore vr: Networkable virtual reality software tools for real-time composition and performance. International conference on new interfaces for musical expression

Chapman O (2009) The icebreaker: Soundscape works as everyday sound art. Organised Sound 14(1):83–88

Cherry E, Latulipe C (2014) Quantifying the creativity support of digital tools through the creativity support index. ACM Trans Comp-Human Interact 21(4):21

Ciciliani M (2020) Virtual 3D environments as composition and performance spaces. J New Music Res 49(1):104–113

Costa W, Ananias L, Barbosa I, Barbosa B, De'Carli A, Barioni R et al. (2019) Songverse: a music-loop authoring tool based on virtual reality. IEEE symposium on virtual and augmented reality (pp. 216–222)

Drever JL (2002) Soundscape composition: the convergence of ethnography and acousmatic music. Organ Sound 7(1):21–27

d'Escriván J (2009) Sound art (?) on/in film. Organ Sound 14(1):65–73

Eckel G (2001) Immersive audio-augmented environments: the listen project. Proceed Fifth Int Conf Infor Visual 128:571–573

Favory X, Fonseca E, Font F, Serra X (2018) Facilitating the manual annotation of sounds when using large taxonomies. Proceedings of the 23rd conference of open innovations association fruct, pp. 60–64

Fonseca E, Pons Puig J, Favory X, Font Corbera F, Bogdanov D, Ferraro A et al. (2017) Freesound datasets: a platform for the creation of open audio datasets. Proceedings of the international society for music information retrieval conference, pp. 486–493

Font F (2021) SOURCE: a Freesound Community Music Sampler. Proceedings of audio mostly conference, pp. 182–187

Font F, Brookes T, Fazekas G, Guerber M, La Burthe A, Plans D et al. (2016) Audio commons: bringing creative commons audio content to the creative industries. Audio engineering society conference: 61st international conference: audio for games

Font F, Roma G, Serra X (2013) Freesound technical demo. Proceedings of the ACM international conference on Multimedia, pp. 411–412

Freeman J, Disalvo C, Nitsche M, Garrett S (2011) Soundscape composition and field recording as a platform for collaborative creativity. Organ Sound 16(3):272–281

Graham R, Cluett S (2016) The soundfield as sound object: Virtual reality environments as a three-dimensional canvas for music composition. Audio engineering society conference: 2016 AES international conference on audio for virtual and augmented reality

Hart SG (2006) NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting, Vol. 50, pp. 904–908

Huang C, Lin Y, Zeng X, Newman M, O'Modhrain S (2015) Olegoru: a soundscape composition tool to enhance imaginative storytelling with tangible objects. Proceedings of the ninth international conference on tangible, embedded, and embodied interaction, pp. 709–714

Koutsomichalis, M. (2013). On soundscapes, phonography and environmental sound art. Journal of sonic studies, 4(1).

Leonard M, Strachan R (2014) More than background: Ambience and sound-design in contemporary art documentary film. Music and sound in documentary film, pp. 180–193. Routledge

Masu R, Bala P, Ahmad M, Do Nascimento Correia N, Nisi V, Nunes N, Romão T (2020) Vr open scores: scores as inspiration for vr scenarios. Proceedings of the international conference on new interfaces for musical expression, pp. 109–114

Moore A, Howell M, Stiles A, Herrera N, McMahan R (2015) Wedge: A musical interface for building and playing composition-appropriate immersive environments. IEEE symposium on 3d user interfaces, pp. 205–206

Polfreman R (2009) Frameworks 3D: composition in the third dimension. Proceedings of the international conference on new interfaces for musical expression, pp. 226–229

Schafer RM (1977) The tuning of the world. Random House Inc

Schirosa M, Janer J, Kersten S, Roma G (2010) A system for soundscape generation, composition and streaming. Xvii cim-colloquium of musical informatics

Skach S, Xambó A, Turchet L, Stolfi A, Stewart R, Barthet M (2018) Embodied interactions with e-textiles and the internet of sounds for performing arts. Proceedings of the international conference on tangible, embedded, and embodied interaction, pp. 80–87. ACM

Thorogood M, Pasquier P (2013) Computationally created soundscapes with audio metaphor. International conference on computational creativity, pp. 1–7

Truax B (1992) Electroacoustic music and the soundscape: the inner and outer world. Comp Contemp Music Thought 1:374–398

Truax B (1996) Soundscape, acoustic communication and environmental sound composition. Contemp Music Rev 15(1–2):49–65

Truax B (2008) Soundscape composition as global music: electroacoustic music as soundscape. Organ Sound 13(2):103–109

Turchet L, Barthet M (2018) Jamming with a smart mandolin and Freesound-based accompaniment. IEEE Conference of open innovations association (FRUCT), pp. 375–381

Turchet L, Fazekas G, Lagrange M, Shokri Ghadikolaei H, Fischione C (2020) The internet of audio things: state-of-the-art, vision and challenges. IEEE Intern Thing J 7(10):10233–10249

Turchet L, Hamilton R, Çamci A (2021) Music in extended realities. IEEE Access 9:15810–15832

Turchet L, Serafin S (2013) Investigating the amplitude of interactive footstep sounds and soundscape reproduction. Appl Acoust 74(4):566–574

Turchet L, Zanetti A (2020) Voice-based interface for accessible soundscape composition: composing soundscapes by vocally querying online sounds repositories. Proceedings of audio mostly conference, pp. 160–167

Turner P, McGregor I, Turner S, Carroll F (2003) Evaluating soundscapes as a means of creating a sense of place. Proceedings of international conference on auditory display, pp. 148–151

Valle A, Armao P, Casu M, Koutsomichalis M (2014) Soda: A sound design accelerator for the automatic generation of soundscapes from an ontologically annotated sound library. International computer music conference

Valle A, Lombardo V, Schirosa M (2009) Simulating the soundscape through an analysis/resynthesis methodology. Auditory display, pp. 330–357. Springer

Verron C, Aramaki M, Kronland-Martinet R, Pallone G (2009) A 3-d immersive synthesizer for environmental sounds. IEEE Trans Audio, Speech and Lang Process 18(6):1550–1561

Westerkamp H (2002) Linking soundscape composition and acoustic ecology. Organ Sound 7(1):51–56

Wrightson K (2000) An introduction to acoustic ecology. Soundscape: The J Acoust Ecol 1(1):10–13

Zappi V, Mazzanti D, Brogni A, Caldwell D (2012) Concatenative synthesis unit navigation and dynamic rearrangement in vrGrains. Proceedings of the sound & music computing conference, pp. 403–408