Mixed Scape: Development of Framework and Artwork for Auditory Correspondence in Mixed Reality

Seungsoon Park **Jongpil Lee** **Taewan Kim**
Neutune Research
{seungsoonpark, jongpillee, taewankim}@neutune.com

Tae Hong Park
NYU Steinhardt School of Culture, Education, and Human Development
thp1@nyu.edu

Joonhyung Bae **Juhan Nam**
Graduate School of Culture Technology, KAIST
{jh.bae, juhan.nam}@kaist.ac.kr

ABSTRACT

In this paper, we introduce mixed scapes which builds on mixed reality (MR) concepts through the lens of a soundscape-focused research framework. We summarize (1) research activities that focus on relationships between environmental sound and various forms of realities, (2) a research framework for auditory responses to MR studies, and (3) artwork examples that demonstrate key concepts outlined in this paper. We also discuss integration of research modules where soundscape studies, musique concrète research, sound analysis, and artificial intelligence are combined into a single framework that, we believe, facilitates exploration, study, and expansion of mixed reality systems. The four installation works that are briefly presented here demonstrate potential pathways for a research-led-practice and practice-led-research model that blends technical research and artistic research within the realm of mixed reality domains.

1. INTRODUCTION

Environmental sound is a multidisciplinary research topic that has attracted researchers and practitioners from various fields including the arts, sciences, social sciences, and engineering. From the perspective of composers, environmental sound affords artists unusually rich materials for inspiration and creative expression. In this scenario, the artist oftentimes negotiates one's role as the composer vs. reporter and balances between composition vs. documentation [1]. In areas of sound design, gaming industry, motion picture industry, as well Foley, the importance of environmental sound - whether synthesized, imitated, recorded and played back verbatim, or a combination of various techniques - cannot be overstated. Similarly, scientific interests, and in particular in the context of human auditory processing and understanding [2], have also paved the way in pushing forward research efforts in auditory processing emulation and automatic analysis of acoustic scene analysis worldwide [3, 4].

In the field of mixed reality (MR) and the notion of mediality continuum that includes augmented reality (AR) and virtual reality (VR), the importance of environmental sound has also solidified as critical in producing realistic virtual acoustic environments [5]. In the case of MR, literature shows that environmental sound research is primarily focused on sound spatialization[6]; and while impression of sound sources within a 3D space is important, we also recognize opportunities in furthering this field as a techno-artistic research proposition. This effort is framed around a term we call *Mixed Scape* and position its various concepts along the spectrum of mixed reality.

To this end, we first provide a general overview of mixed reality with focus on environmental sound. We then discuss environmental sound and its application in sound art including musique concrète, "noise" music, soundscapes, and sound design. Lastly, we review recent advancements in environmental sound synthesis that leverage machine learning and cloud computing technologies.

In addition, we will formalize a research framework of *Mixed Scape* and introduce four artworks that implement the framework. These works were presented at physical galleries and virtual online exhibition spaces. Through this effort, we explore how environmental sound can help mediate artistic and technological creativity in MR.

2. RELATED RESEARCH

2.1 Mixed Reality: An Auditory Perspective

2.1.1 Mixed, Mediated and Multimediated Reality

The concept of mixed reality (MR) can be thought of as a continuum between reality and virtuality [7] where it is composed of reality, augmented reality (AR), augmented virtuality, and virtual reality (VR). Virtual reality, arguably most familiar to the public, involves a simulated world that is outside of the real world and generates real experiences through digital means. AR, on the hand, can be defined as a world where computer-generated content is blended with the real world. Mann further extended the concept of MR, and proposed the concept of mediated reality, in which reality is mixed, blended, and modified [8]. His group also devised the concept of multimediated reality and the notion of multidimensional, multimodal, and multisensory realities [9].

2.1.2 Audio Augmented Reality (AAR)

Audio augmented reality (AAR) is the overlaying of physical worlds with virtual audio layers that have associated information: e.g. visitors hearing geo-tagged descriptions of paintings in museums or art galleries as they walk from room to room or tilt their heads. McGill et al. further discuss positive impacts that acoustic clarity can bring to an individual's sonic perception using headsets by correlating the idea of auditory mixed reality [6]. Examples include an automatic voice-guided tour experience [10] and user-adaptive museum audio guiding systems [11].

2.1.3 3D Sound in Virtual Environments

3D spatial sound perception is one of the most robust research fields in VR, where according to Mereu, a 3D sound environment improves the "feeling" of space in virtual worl- ds. Common topics and spatialization methods in this research area include distance, angles, and 3D positions of listener and sound objects [12, 13]. While a number of different types of 3D sound tools exist, the Unity VR SDK is perhaps one of the more widely used software packages [14]. Unity allows user-customization of various sound features, such as location, movement, types of spaces within virtual environments. Another example is Spatialisateur(SPAT), which enables control over a sound object's position in a 3D space adapted to different viewpoints [15].

From the examples above, we observe that AAR emphasizes a user-centric acoustic environment control approach addressing immersive audience experience. In 3D sound, however, the focus is more geared towards designing virtual environments. Our proposed approach attempts to build on AR and VR research to facilitate integration, harnessing, and connection within mediality continua, while also emphasizing artists' perspectives in imagining and creating environmental sounds and soundscapes.

2.2 Environmental Sound: Art

Since the advent of recording technologies, artists have been engaged in exploring a diverse spectrum of music-making using sounds that were traditionally considered non-musical and noisy. Machines, devices, and technologies such as the Victrola, magnetic tape, handheld recording devices, and smartphones have opened up musical possibilities and expanded musical genres where environmental sounds oftentimes can play a critical role. In this section we briefly summarize environmental sound-inspired examples.

2.2.1 The Art of Noises and Musique Concrète

An early example of mechanical devices for musical expression was devised by the futurist Luigi Russolo who is largely credited for developing the *Intonarumori*. While this instrument did not record or reproduce recorded sounds, it did contribute in introducing futurism aesthetics of technology, youth, and machines such as the cars, airplanes, and industrialized cities that were "recreated" in concert music settings [16].

Pierre Schaeffer is widely regarded as the father of *musique concrète* who proposed the "sound object" concept or a sound that is devoid of meaning, and thus, enabling attribution of new meanings [17]. In a series of etudes introduced in 1948, he used locomotive sounds as source material for a new style of musical composition referred to as *musique concrète*. In particular, he used familiar train sounds applying de-contextualisation methods, sonic deviation and abstract rhythm arrangement techniques *Etudes aux chemins de fer* while also devising various sound manipulation methods including pitch shifting, lock-groove record looping, sound excerpting, and filtering. These techniques would later become basic tools in found sound composition.

2.2.2 Soundscapes: Acoustic Ecology

The term soundscapes is perhaps most commonly associated by the World Soundscape Project (WSP) founded by R. Murray Schafer and launched in the late 1960s as an effort to draw attention to the rise of noise pollution in Canada [18]. The project's primary focus was on recording and archiving local soundscapes as a means of observation and preservation. While the recordings were generally "unedited" onsite recordings, some examples also showed beginnings of more artistic intentions utilized in soundscape composition, whereby compressing long recordings into a shorter, montaged "edited" version of the original recording.

2.2.3 Soundscapes: Composition

While WSP focused on soundscape documentation and archiving from an acoustic ecological angle, soundscape composition began to take shape as a compositional style where aspects of both preservation and modulation were embraced.
One of early works in this domain is Ferrari's *Presque Rien* (1970). This work, as the title implies, employed minimal manipulation similar to techniques used in WSP's *Entrance to the Harbour* (1973), highlighting segments of the larger original unedited recording compressing it into a shorter version. Another classic soundscape composition is Truax's *Riverrun* (1986). What makes this piece unique is in creating a virtual soundscape by using granular synthesis techniques with pure sine-wave grains. One can also argue that Edgard Varèse's *Poème électronique* [19] foreshadowed multimodal soundscape composition style as a multimedia production combining soundscape preservation, modulation, sound synthesis, multi-channel sound projection, story-telling, sound diffusion, and early ideas of "cinema for the ears." This style would later be seen in compositions that utilized elements of soundscape composition by composers such as Appleton (Times Square Times Ten and San Francisco Airport Rock), Smalley (Empty Vessels and Valley Flow), and Park (Aboji/Omoni and 48 13 N, 16 20 O).

2.2.4 Sound Design

In the area of cinema, television, and media, environmental sounds have been used to emphasize and augment sonic atmospheres and scenes. Interestingly enough, however, to make sounds more realistic, Foley artists are oftentimes relied upon to re-create sounds through entirely unrelated means. For example, in the case of bird wing flaps a well-known trick is to use pair of gloves; and in other cases,

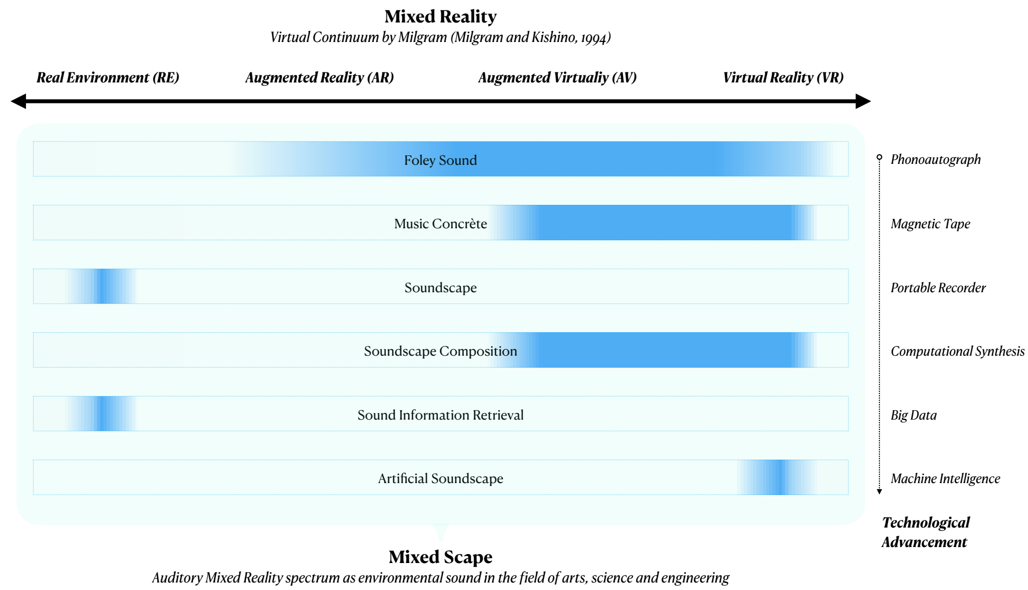

Figure 1. Environmental Sounds in Mixed Reality

some sounds are rendered entirely through synthetic means [20]. Regardless of Foley techniques used, however, we observe that the resulting sound effects maintain properties of augmented reality.

The above categories and associated examples demonstrate that environmental sounds are dealt with in numerous ways for both artistic and sound design purposes. The following sections describe examples of environmental sounds from the point of view of scientific and engineering research including computer-based sound analysis and sound synthesis.

2.3 Technical Advances in Environmental Sound Research

Recent technological advances in machine learning, cloud computing, and the Internet of Things (IoT) has led to opportunities where environmental sound can be captured, classified, quantified, stored, and easily retrieved remotely through the Internet. In this section, we review a number of such studies.

2.3.1 Acoustic Scenes and Events Analysis

An initial research path in the quest to better understand soundscapes, and in particular acoustic scenes, has led to efforts of dataset creation and development of machine learning techniques. In the case of developing soundscape datasets, a combination of crowd-sourcing techniques and using existing sound archives (e.g. freesound.com) was employed where researchers collected, organized, labeled, and edited raw audio data rendering ready-to-use environmental sound datasets and associated metadata [21, 22]. Subsequent research paths focused efforts on soundscape-based machine learning. In particular, workshops such as Detection and Classification of Acoustic Scenes and Events (DCASE) have gained popularity in recent years, in part, due to its application potential in areas such as autonomous driving and home appliances [4]. Additionally, the concept of real-time soundmaps, sound sensor networks, automatic acoustic event detection, and urban noise have also developed alongside and in parallel where citygram (2011) introduced a so-called 3D - data-Driven, community-Driven, art-Driven model [23, 24, 25].

2.3.2 Artificial Soundscape Synthesis

A common approach that can be seen in the area of soundscape synthesis is the use of metadata. For example, Cano et al. utilized a semi-automatic system for generating ambient sounds where the input would entail text queries and the output audio stems related to the input queries [26]. Birchfield et al. developed a generative model [27] using WordNet that linked semantic information associated with individual sound files and automatic sound uploading tool based on JSyn. Thorogood and Pasquier designed a soundscape composition engine called Audio Metaphor [28]. This system integrates a search method that semantically relates sound data in an online database by gathering natural language queries using the Twitter API. A final example is Soundscaper by Park and Murakami [29] employing a synthesis-by-analysis soundscape synthesis approach. Soundscaper generates artificial soundscapes through content-based synthesis algorithms which allows real-time control over arbitrary duration, density, and “loudness” beyond simply linear scaling of resulting waveform using concepts foreground, middle-ground, and background soundscape dimensions.

3. MIXED SCAPE

In the preceding sections, we review a number of examples and topics in environmental sound research that contribute to the auditory spectrum of MR from both artistic and technical research domains. From this perspective, we propose the concept *Mixed Scape*.

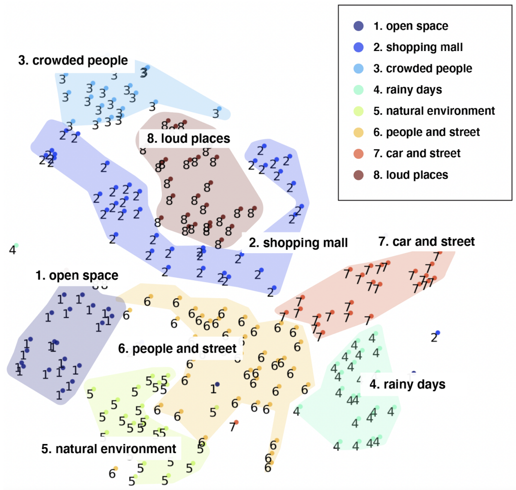

Figure 2. *Eight Soundscape Templates*

3.1 Mixed Scape Framework

The concept of Mixed Scape is perhaps best articulated by considering it as characteristics of virtuality, reality, and augmentation through the lens of four main components as shown in Figure 1: (1) music concrete and "noise" music, (2) soundscape and acoustic ecology, (3) soundscape and composition, and (4) sound design. (1) corresponds to an augmented virtual environment created through the process of environmental sound abstraction and de-contextualization as intended by the composer. (2) on the other hand, can be regarded as being related to a "real" environment as it corresponds to the original, unaltered environment sound recording and is absent of augmentation. In (3), both preservation and modulation of environmental sound are characteristic in rendering virtual and artificial outputs that can be further augmented. Lastly, in (4), resulting sounds are typically rendered as part of a medium, narrative, or scene. In particular, when used to reinforce raw field-recordings, it has characteristics of augmented reality; and when used to create imagined, synthetically generated sounds, it gravitates towards virtual reality.

3.2 Mixed Scape Artwork Examples

In this section we report on four mixed scape artworks representing various sound characteristics within a mixed reality sound continuum. Four methodologies are used to four each of the artwork presented including *Technical Research*, *Sound Archiving*, *Audio-Visual Medium*, and *Installation Art*. The final artwork was exhibited both in a physical gallery space and in a remotely accessible virtual space.

3.2.1 Eight Soundscape Templates

As we go about our daily lives, the soundscapes that surround us change as we go from place to place and room to room. In some cases, the changes are drastic - e.g. exiting a quiet building and onto a busy city center street - while in other instances, the changes are more gradual and subtle - e.g. traversing a large park. Postulating that is a finite number of soundscapes types or "sound carpet" prototypes similar in concept to musical dynamics (e.g. *pp, p, mp, mf, f, ff*) that are applicable to soundscapes [29], we attempted to computationally test this hypothesis on 229 hours of raw urban video recordings [1]. The first step entailed feature extraction using a pre-trained audio classification model based deep neural networks [30]. The second step was subject to K-means clustering of the features where the number of centroids were determined using silhouette analysis. In the final step, eight main clusters were selected by manually viewing and listening to the video and its associated soundscapes. The resulting eight clusters or "sound carpets" included *car and street*, *shopping mall*, *open space*, *people and street*, *crowded people*, *rainy days*, *natural environment*, and *loud places*. Using these categories, we created an artwork called *Eight Soundscape Templates* as shown in Figure 2. One interesting finding was that videos in the *rainy days* cluster included not only scenes and sounds that had rain but also beach-side scenes similar to what one could imagine in Foley-based sound design situations. While further analysis, more data, and testing is needed, initial findings show the potential for contextual association of real world visual dimensions with real-world auditory dimensions. Furthermore, when using environmental sound recordings as a source for musical material, we believe that new compositional possibilities will emerge when soundscape database navigation and selection tools are developed to organize the vastness of soundscape recordings.

Figure 3. *How does Foley sound acquire AR characteristics?*

3.2.2 How does Foley Sound Acquire AR Characteristics

The main goals of this artwork was to demonstrate (1) Foley sound processes generating AR characteristics, (2) impact of (1) on perceived reality, and (3) visualize and display (1) and (2) as a cultural interface. In particular, we note that When Foley sound acquires AR characteristics, there are two distinct phases. The first occurs when the audio associated within a scene of a video object is augmented at the post-production stage. When the main-production is terminated by satisfying certain conditions, the video acquires actualized virtuality. Actualization is that virtuality becomes actuality through complex events [31]. Therefore, Foley work becomes the act of giving AR properties. The second is the moment when a costume player recreates the Foley sound that enhances the sensory and conceptual quality of the costume play. In the process of

[1] https://www.youtube.com/channel/UCSZXwdBppzhYpcqf69WNOIg

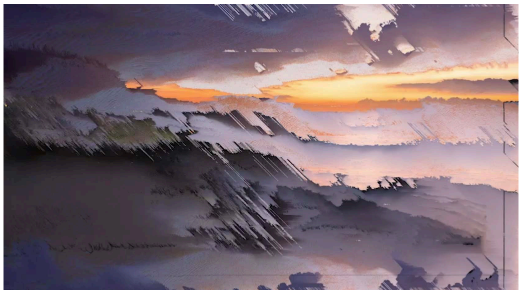

Figure 4. *Audio Visual for Concrete Music*

imitating a character, costume players embody individual desires and interact with others. For this reason, the costume player tries to implement the costume and Foley sound in high quality to actualize the cosplay. To explore this, we created an "archival" video artwork using a variety of Foley sounds extracted from films and animations, and presented the work in the form of a roll screen poster as shown in Figure 3. This work further highlighted the human auditory system's malleability in accepting new realities rendered through augmented and virtual environmental sounds.

3.2.3 Hyper Landscape

One way to view the transformation of unedited environmental recordings as a basis for music composition is through an augmented virtuality perspective that corresponds to musique concrète. To illustrate this idea, we designed a "hyper pixel layout" artwork to visually express structural arrangements and transformations of real objects on a virtual environment. As "real environments" and their realism can often be intuitively grasped, we attempted to create a visual "differential phenomenon" from the perspective of inter-reality so that clear recognizably real worlds would be made more ambiguous. First, we selected a pair of image and sound corresponding to the 'eight soundscape template' above. It then maps the frequency band of the sound to the RGB value of the image to select specific pixels that respond to the sound. Detecting the playback of the sound, pixels started to scatter in random directions, obscuring the perception of the image. This leads us to perceive the 'real world' revealed in the image as a 'virtual space' based on reality. This is articulated in *Hyper Landscape* as shown in Figure 4 where meanings that float and change are captured and presented visually using touch designer and the processing software.

3.2.4 Mixed Artificial and Virtual Soundscape

Neuroscape [32] is a system that automatically searches for soundscape using images as input where, for example, a forest image returns associated sounds such as wind and bird chirping sounds. The system uses deep learning techniques to first analyze input images to detect scenes and objects via Google's Cloud Vision API [2]. Words linked to detected scenes and objects then are linked to 527 audio categories [21] in a dense Word2Vec space [33]. In the final step, sounds are ranked and retrieved from the Neuroscape database using an audio classification model. These sounds are then used as source material for the artificial soundscape rendering [30] as illustrated in Figure 5.

[2] https://cloud.google.com/vision

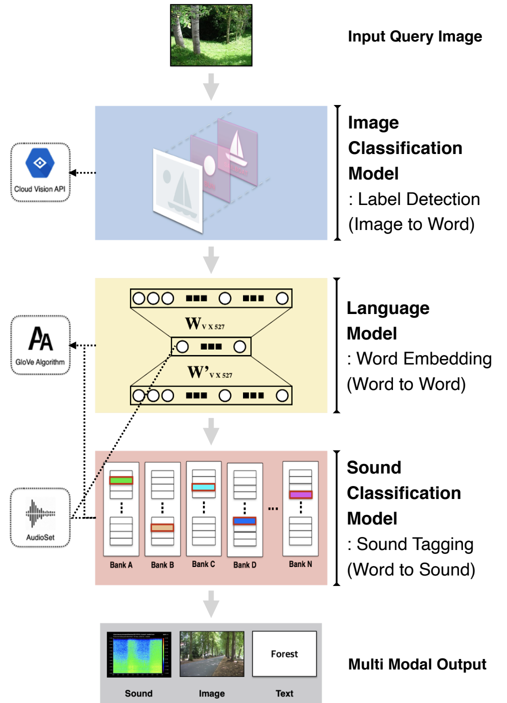

Figure 5. The Neuroscape system

Using the Neuroscape system we created an interactive installation artwork where off-the-shelf MIDI fader controllers were used to facilitate audience participation without requiring musical or technical expertise. We devised Steve Mann's multimediated reality concept to render the artwork [9] as shown in Figure 6. Here, we see audiences' fader left/right interaction where hard left results in pure artificial soundscape and hard right results in virtual soundscape generation. While the fader created user-generated artificial and virtual soundscapes, the audience was also able to experience a mixed soundscape. The various soundscapes were in many cases, were entirely synthesized as urban, natural, and other types of landscape videos oftentimes did not include associated sounds: we generated artificial soundscapes using the Neuroscape system, and used the synthesized soundscapes to link videos lacking audio. Accordingly, we combined and mapped computer-generated soundscapes to several video clips to render audio visual effects using VDMX.

The Neuroscape installation's primary goal was to situate humans within multidimensional landscapes through visualization of real, artificial, and virtual audio-visual landscapes. Furthermore, as artificial and virtual landscape sounds are sometimes perceived as being more realistic than recorded and unedited counterparts, audiences were indirectly asked to negotiate and distinguish between real and virtual sound environments.

3.2.5 Mixed Scape XR Exhibition

Due to the unexpected complications brought about by

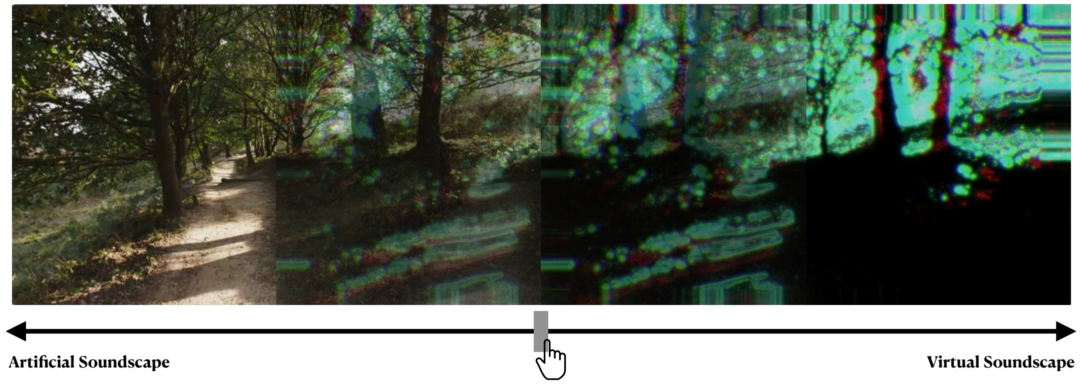

Figure 6. *Mixed Artificial and Virtual Soundscape*

COVID-19, in-person participation of exhibitions was quite limited. To enable meaningful audience engagement, however, we created a virtual exhibition space using Mozilla hub where audiences were able to "walk" through the XR exhibition space via keyboard controls using a standard web-browser. The virtual exhibition space, while not ideal, offered an alternative way for audiences to safely enjoy installation artworks with the benefit of 24/7 access from the comfort of one's home.

4. CONCLUSIONS

In this paper, we presented *Mixed Scape*, a framework for research and artistic exploration based on environmental sound within Mixed Reality domains. We redefined the spectrum of auditory mixed realities by reviewing existing continua concepts of MR and exploring artistic and scientific research examples in environmental sound. In addition, we examples of four *Mixed Scape* artworks. We presented *Mixed Scape* as an indication of research and artistic creation dealing with the reality of sound within AR, VR, MR, and XR. We aim to further supplement ideas described in this paper and will improve and explore the possibility of live performances in online and offline environments.

Acknowledgments

This project was supported by the Transdisciplinary Research (TDR) Program at Korea Institute for Advanced Study (KIAS). We would like to express our gratitude to Prof. Sungju Woo for her insightful comments and suggestions.

5. REFERENCES

[1] T. H. Park, "Composition vs. Documentation," 2008. [Online]. Available: https://cec.sonus.ca/events/TES/2008/park_friction_abstract.html

[2] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT Press, 1994.

[3] R. F. Lyon, *Human and Machine Hearing: Extracting Meaning from Sound*. Cambridge University Press, 2017.

[4] A. Mesaros, T. Heittola, A. Diment, B. Elizalde, A. Shah, E. Vincent, B. Raj, and T. Virtanen, "DCASE 2017 challenge setup: Tasks, datasets and baseline system," in *Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2017.

[5] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl, "Sonic interactions in virtual reality: state of the art, current challenges, and future directions," *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 31–43, 2018.

[6] M. McGill, S. Brewster, D. McGookin, and G. Wilson, "Acoustic transparency and the changing soundscape of auditory mixed reality," in *CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–16.

[7] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Transactions on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[8] S. Mann, "Mediated Reality," M.I.T. Media Lab Perceptual Computing Section, Cambridge, Massachusetts, http://wearcam.org/mr.htm, TR 260, 1994.

[9] S. Mann, T. Furness, Y. Yuan, J. Iorio, and Z. Wang, "All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality," *arXiv preprint arXiv:1804.08386*, 2018.

[10] A. Zimmermann and A. Lorenz, "LISTEN: a user-adaptive audio-augmented museum guide," *User Modeling and User-Adapted Interaction*, vol. 18, no. 5, pp. 389–416, 2008.

[11] L. Cliffe, J. Mansell, C. Greenhalgh, and A. Hazzard, "Materialising contexts: virtual soundscapes for real-world exploration," *Personal and Ubiquitous Computing*, pp. 1–14, 2020.

[12] D. R. Begault and L. J. Trejo, "3-D sound for virtual reality and multimedia," 2000.

[13] S. W. Mereu, "Improving depth perception in 3 D interfaces with sound," Master's thesis, Citeseer, 1995.

[14] A. Çamcı, P. Murray, and A. G. Forbes, *A Web-based System for Designing Interactive Virtual Soundscapes*.

Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2016.

[15] J.-M. Jot and O. Warusfel, "Spat: A spatial processor for musicians and sound engineers," in *CIARM: International Conference on Acoustics and Musical Research*, 1995.

[16] B. Brown, "The noise instruments of Luigi Russolo," *Perspectives of New Music*, pp. 31–48, 1981.

[17] C. Palombini, "Machine Songs V: Pierre Schaeffer: From research into noises to experimental music," *Computer Music Journal*, vol. 17, no. 3, pp. 14–19, 1993.

[18] R. M. Schafer, *The soundscape: Our sonic environment and the tuning of the world.* Simon and Schuster, 1993.

[19] O. Mattis, "Varèse's Multimedia Conception of" Déserts"," *Musical Quarterly*, pp. 557–583, 1992.

[20] V. Dakic, "Sound design for film and television," 2009.

[21] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 776–780.

[22] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in *ACM International Conference on Multimedia*, 2014, pp. 1041–1044.

[23] T. H. Park, J. Turner, M. Musick, J. H. Lee, C. Jacoby, C. Mydlarz, and J. Salamon, "Sensing Urban Soundscapes." in *EDBT/ICDT Workshops*, 2014, pp. 375–382.

[24] B. Zerza and T. Park, "The City of the Future: The Urban (Un) Seen Connecting Citizens and Spaces via Community Sensing," in *IOP Conference Series: Earth and Environmental Science*, vol. 588, no. 3, 2020, p. 032011.

[25] J. P. Bello, C. Silva, O. Nov, R. L. Dubois, A. Arora, J. Salamon, C. Mydlarz, and H. Doraiswamy, "Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution," *Communications of the ACM*, vol. 62, no. 2, pp. 68–77, 2019.

[26] P. Cano, L. Fabig, F. Gouyon, M. Koppenberger, A. Loscos, and Á. Barbosa, "Semi-Automatic Ambiance Genereation," 2004.

[27] D. Birchfield, N. Mattar, and H. Sundaram, "Design of a generative model for soundscape creation," in *International Computer Music Conference (ICMC)*, 2005.

[28] M. Thorogood, P. Pasquier, and A. Eigenfeldt, "Audio metaphor: Audio information retrieval for soundscape composition," *Sound and Music Computing Conference (SMC)*, pp. 277–283, 2012.

[29] T. H. Park and S. Murakami, "The Soundscaper: A Tool for Soundscape Re-Synthesis," 2020.

[30] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold *et al.*, "CNN architectures for large-scale audio classification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 131–135.

[31] P. Lévy, *Qu'est-ce que le virtuel?* Éditions La Découverte, 1995.

[32] S. Park, J. Lee, and J. Nam, "NEUROSCAPE: artificial soundscape based on multimodal connections of deep neural networks," in *International Computer Music Conference (ICMC) (Installation)*, 2018.

[33] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.