

Review

Spatial Audio for Soundscape Design: Recording and Reproduction

Joo Young Hong ^{1,*}, Jianjun He ², Bhan Lam ¹, Rishabh Gupta ¹ and Woon-Seng Gan ¹

¹ School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore; blam002@e.ntu.edu.sg (B.L.); grishabh@ntu.edu.sg (R.G.); ewsgan@ntu.edu.sg (W.-S.G.)

² Maxim Integrated Products Inc., San Jose, CA 95129, USA; jianjun.he@maximintegrated.com

* Correspondence: jyhong@ntu.edu.sg; Tel.: +65-8429-7512

Academic Editor: Gino Iannace

Received: 29 March 2017; Accepted: 9 June 2017; Published: 16 June 2017

Featured Application: This review introduces the concept of spatial audio in the perspective of soundscape practitioners. A selection guide based on the spatial fidelity and degree of perceptual accuracy of the mentioned spatial audio recording and reproduction techniques is also provided.

Abstract: With the advancement of spatial audio technologies, in both recording and reproduction, we are seeing more applications that incorporate 3D sound to create an immersive aural experience. Soundscape design and evaluation for urban planning can now tap into the extensive spatial audio tools for sound capture and 3D sound rendering over headphones and speaker arrays. In this paper, we outline a list of available state-of-the-art spatial audio recording techniques and devices, spatial audio physical and perceptual reproduction techniques, emerging spatial audio techniques for virtual and augmented reality, followed by a discussion on the degree of perceptual accuracy of recording and reproduction techniques in representing the acoustic environment.

Keywords: soundscape; spatial audio; recording; reproduction; virtual reality; augmented reality

1. Introduction

Urban acoustic environments consist of multiple types of sound sources (e.g., traffic sounds, biological sounds, geophysical sounds, human sounds) [1]. The different types of sound sources have different acoustical characteristics, meanings and values [2–4]. Conventionally, environmental noise policies are primarily centered on the energetic reduction of sound pressure levels (SPL). However, SPL indicators provide limited information on perceived acoustic comfort as it involves higher cognitive processes. To address the limitations of traditional noise management, the notion of a soundscape has been applied as a new paradigm. Schafer, a Canadian composer and music educator, introduced the term soundscape to encompass a holistic acoustic environment as a macrocosmic musical composition [5]. In this context, soundscape considers sound as a resource rather than waste and focuses on people’s contextual perception of the acoustic environment [6].

Due to rising prominence of the soundscape approach, ISO TC43 SC1 WG 54 was started with the aim of standardizing the perceptual assessment of soundscapes. As defined in ISO 12913-1: definition and conceptual framework, acoustic environment is “sound from all sound sources as modified by the environment” and the modification by the environment includes effects of various physical factors (e.g., meteorological conditions, absorption, diffraction, reverberation, and reflection) on sound propagation. This implies that soundscape is a perceptual construct related to physical acoustic environment in a place.

According to ISO 12913-1, the context plays a critical role in the perception of a soundscape, as it affects the auditory sensation, the interpretation of auditory sensation, and the responses to the acoustic environment [7]. The context includes all other non-acoustic components of the place (e.g., physical as well as previous experience of the individual). Herranz-Pascual et al. [8] proposed a people-activity-place framework for contexts based on four clusters: person, place, person-place interaction and activity. The suggested framework shows interrelationships between person and activity and place, which may influence a person’s experience of the acoustic environment.

It is evident from [8] that the human auditory process is deeply entwined in the person-activity-place framework. Thus, sufficient insight about the human auditory process is imperative to record and reproduce an acoustic environment with sufficient perceptual accuracy for proper analysis. For simplicity, Zwicker and Fastl has grouped the human auditory processing system into two stages: (1) the preprocessing in the peripheral system, and (2) information processing in the auditory system [9].

Current reproduction methods (e.g., loudspeakers, headphones, etc.) have sufficient fidelity for proper interpretation of frequency and temporal characteristics in the peripheral system (stage 1). As the spatial dimension of sound is only interpreted in stage 2, where there is significant exchange of information between both ears, complexity of reproduction is greatly increased. For instance, to locate a sound source in space, the characteristics of intensity, phase, and latency must be presented accurately to both ears. Hence, spatial audio recording and reproduction techniques should be reviewed in the technological perspective with a focus on their relationship with soundscape perception—the main goal of this paper.

As recording and reproduction techniques are heavily employed in the soundscape design process, a brief discussion will shed light on the areas where such techniques are most commonly used. The soundscape design process can be summarized into three stages, as illustrated in Figure 1 [10]. Stage 1 aims to define and analyze existing soundscapes. In this stage, soundscape researchers and planners are required to evaluate and identify ‘wanted sounds’ to be preserved or added, and ‘unwanted sounds’ to be removed or reduced based on the context. In Stage 2, soundscape planning and design scenarios are proposed based on the analysis of the existing soundscapes. The soundscape design proposals will then be simulated to be objectively and subjectively evaluated by stakeholders to determine the final soundscape design. In Stage 3, the final soundscape design will be implemented in situ. After implementation, validation of the soundscape design will be performed with iteration of Stage 1 for analysis of implemented soundscape design.

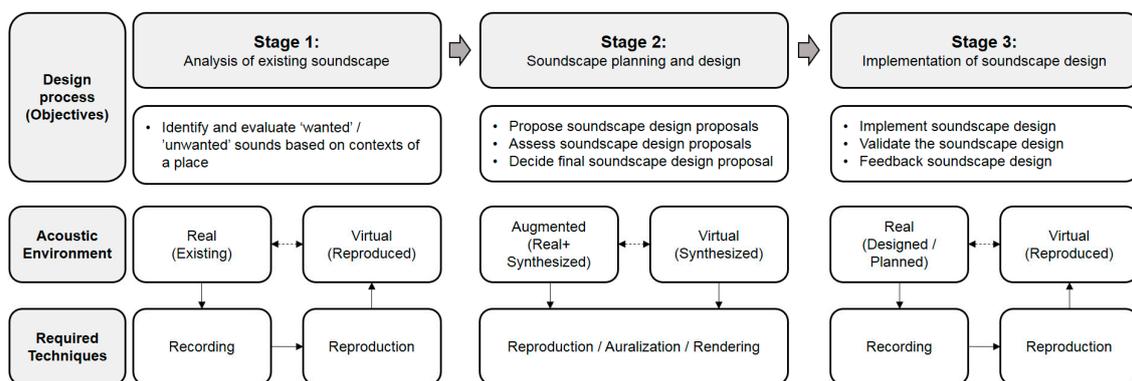


Figure 1. Schematic illustration of soundscape design process, types of acoustic environment and required techniques.

Throughout the soundscape design process, soundscapes can be evaluated based on the real acoustic environment or virtual (reproduced or synthesized) environment [6,11]. Similarly, in Stage 1 (analysis of existing soundscape) and Stage 3 (implementation of soundscape design), soundscapes

can be assessed in real or reproduced acoustic environments. The reproduced acoustic environment is constructed from audio recordings in the real world. If the physical location for soundscape implementation is not available (i.e., not built), future soundscape design scenarios might also be evaluated using synthesized (or simulated) acoustic environment in Stage 2 (soundscape planning and design). Moreover, augmented acoustic technologies consisting of synthesized and real sounds can also be applied in soundscape design stage if the physical locations are available.

A real acoustic environment is evaluated *in situ* by means of field study methods, such as soundwalk and behavior observations, which generally focus on the short-term *in situ* experience [6,11]. Narrative interview and social survey usually deals with long-term responses or recalled experiences based on the real acoustic environment [12–14]. In contrast, the virtual acoustic environment, created by recording followed by reproduction, are usually evaluated in a laboratory conditions [6,11]. Aletta et al. [11] reviewed the advantages and limitations of such methods for soundscape assessment. Evaluation methods in the real acoustic environments are advantageous as they provide the most realistic representation of the real-life settings, which can guarantee high ecological validity. However, those methods are subjected to uncontrolled factors, such as temperature, humidity, daylight and wind speed. Accordingly, such methods may yield results that are difficult to generalized as they are limited to a specific context. These methods have been primarily applied to collect and characterize the existing acoustic environment at the early stages in urban soundscape planning and to validate the soundscape design at the implementation stages [6,10,15].

On the other hand, the virtual acoustic environment, are usually evaluated under controlled laboratory conditions. Researchers can design and control the experimental variables, such that it enables us to investigate various casual relationships or correlations. A laboratory-based experiment can obtain results that minimize the effects of other distracting environmental factors. Regardless of the strengths of the laboratory experiments based on virtual acoustic environment, criticisms on discrepancies between virtual and real environments, such as the absence of physical contact and possible perceptual alterations, have been raised. In general, the more control researchers exert in a laboratory experiment, the less ecological validity the findings have, which can be less generalized to real-life scenarios [11,16]. Despite this limitation, acoustic simulation and reproduction techniques can provide powerful tools to evaluate the performance of different soundscape design approaches in Stage 2 before they are implemented in the real world [11].

Currently, the ISO WG54 is working on the Part 2 of the soundscape assessment standard to provide information on minimum reporting requirement for soundscape studies and applications [17]. The reporting requirement includes physical (recording and reproduction techniques) and perceptual data collections both on-site (e.g., soundwalk) and off-site (e.g., laboratory experiments). In particular, recording and reproducing techniques play a critical role to achieve high ecological validity of soundscape research conducted in the virtual acoustic environment. This is because soundscape is the human perception of the reproduced acoustic environment, and the perception is dependent on the perceptual accuracy of the recording and reproduction techniques. Perceptual accuracy is the basis of auralization [18], and is largely influenced by the spatial properties of sounds such as source distance, source location, and reverberation [19,20]. Several laboratory experiments on ecological validation of the reproduction systems have been conducted based on the psycholinguistic measure, such as semantic differential method [16,21,22]. Those studies compared the subjective responses of verbal descriptors on soundscape among *in situ* acoustic environments and the reproduced acoustic environments created by different techniques in laboratory conditions to explore the ecological validity of reproduction systems.

Recently, virtual reality (VR) [23–25] and augmented reality (AR) [26–28] technologies have been increasingly adopted in soundscape studies due to their potential in creating a perceptually accurate audio-visual scene. To create virtual acoustic environments with high ecological validity, it is essential to have a holistic understanding of recording and reproducing techniques with respect to soundscape evaluation methodologies. Even though more studies on the standardization of soundscape recording

and reproduction in an ecologically valid way that reflect context of real acoustic environment are necessary, relatively few studies have been done on these technical aspects of soundscape research. This paper attempts to formulate a framework for the appropriate use of audio recording and reproduction techniques in soundscape design. Sections 2 and 3 systemically reviews the acoustic recording and reproduction techniques that can be adopted in soundscape research, respectively. In Section 4, application of spatial audio in virtual and augmented Reality for soundscape design is discussed. Lastly, the degree of perceptual accuracy of the recording and reproduction techniques in representing acoustic environment is addressed in the discussion.

2. Spatial Audio Recording Techniques for Soundscape Design

Recording of the acoustic environment builds the foundation in soundscape research through analyzing and understanding the soundscape in a real or virtual environment. The recordings, therefore, requires sufficient characteristics of the acoustic environment to be captured for perceptual accuracy. To achieve this, it is necessary to consider two aspects of recording: timbre and spatial qualities. In general, timbre qualities of the recordings are largely dependent on the electrical and acoustic properties of the microphones, such as frequency response, directionality, impedance, sensitivity, equivalent noise level, total harmonic distortion, maximal SPL, etc. [29]. In terms of the spatial aspects, it is common to use multiple microphones for capturing better spatial characteristics of the acoustic environment. In this section, we will review the audio recording techniques, and discuss how they can be applied in the field of soundscape. A comparison of the advantages and disadvantages of these recording techniques are also presented.

2.1. Recording Techniques

In this subsection, we review various recording techniques that are commonly used in audio and acoustics in academic studies and commercial applications.

2.1.1. Stereo and Surround Recording

Sound recording can be conducted in the form of one, two or any number in microphone array. Compared to a single microphone recording, stereo recordings could provide more spatial information on the sound field, including sense of direction, distance, ensemble of the sound stage and ambience. There are basically four types of stereo recording configurations: coincident pair, spaced pair, near-coincident pair, and baffled pair. With the evolution of sound playback system from stereo to surround sound, it is straightforward to record the sound field in surround format, even though most of the surround sound comes from mixing process. Most surround recordings are to be played (directly or after some mixing) via 5.1/7.1 surround sound system. There are various types of surround sound microphone settings. Examples of such surround microphone system include Optimized Cardioid Triangle (OCT) surround, which employs five cardioid mics with one facing front, two facing the sides, and two facing rear; IRT surround, which employs four cardioid mics placed in a square with 90 degree angles; Hamasaki square, which employs four bidirectional mics facing the sides; and a few other configurations. For more practical guidelines on microphone recordings please refer to [30,31].

2.1.2. Microphone Array Recording

To better capture the spatial sound field, more microphones or an array of microphones are required. This idea dated back to the 1930s, where Steinberg and Snow introduced an “acoustic curtain” system that consists of a wall of microphones, which are directly connected to a matching loudspeaker array on the playback end [32]. With a good matching (closely positioned) between the microphone array and loudspeaker array, the recreated sound field is more realistic.

However, such well-matched recording and playback system is hardly practical in today’s applications, where the configurations of the playback system (including headphone, number and

positions of loudspeakers) vary greatly in different room settings. As a result, microphone recordings are no longer used directly without processing, but require real-time or post-processing to analyze the characteristics of the sound field. In general, a microphone array with signal processing could provide three types of the information about the sound field [33]: (i) the locations of the sound sources; (ii) the corresponding (“clean”) sound source signal; and (iii) the characteristics of the sound environment, including the reflections and ambience. Therefore, microphone arrays are widely used in sound source localization, noise reduction, speech enhancement, echo cancellation, source separation, room impulse response estimation, and sound reproduction [34].

Microphone arrays are classified into two categories based on whether they are moving or stationary during the recording. Moving microphone arrays could be used for static or moving sound events. For static sound events, the movements of the microphone array create multiple virtual arrays or enlarges the array size so that more precise information on the sound field could be obtained. A facility enabling the regular movements of the array equipment (e.g., a motor) is required for the moving arrays system. In this case, the assumption that the sound field is spatially static, is very critical. In addition, the effect of recording system movement on the recordings (including the time difference, Doppler effect) should be accounted for. Furthermore, if the movement pattern of the array could be recorded, moving arrays could also be useful in recording non-static sound events. For example, the array could follow the main sound event to give it more focus.

The other category of microphone arrays is the static array. For practical reasons, static arrays are widely adopted as it is easier to set up a static microphone array for recording than moving arrays. Since microphones are to simulate human ears, it is reasonable to establish static microphone arrays as sound reproduction normally assumes the listeners to be static. The most commonly-used static microphone array is the uniform linear array (ULA), which is regularly deployed in beamforming applications. Multiple linear arrays can be arranged in a plane to realize a planar array. Circular array is another commonly used geometric configuration of static microphone array. Compared to linear arrays, circular array provides a two-dimensional scan (360° in the horizontal plane) with uniform performance. Extensions to the circular array include multiple circular arrays at the same plane with different radius, or multiple circular arrays at different planes to form a sphere (i.e., spherical microphone array). Spherical microphone arrays are attractive as they can be decomposed using spherical harmonics, which facilitates the sound field representation and interpretation. Spherical arrays with a large radius can be configured to surround the sound event, creating a third-person perspective, while compact spherical arrays are used to record sound events from the listener’s perspective (i.e., first-person perspective). Theoretical foundations of spherical array processing are discussed in-depth in [35]. As shown in Figure 2, examples of circular or spherical array include B&K [36], Nokia OZO [37], MTB [38], Eigenmike microphone from MHAcoustics [39], and the VisiSonics [40] microphone array. An illustration of various microphone array geometries can be found in [41].



Figure 2. Examples of spherical microphone array (from left to right) from Nokia OZO with 8 microphones, Dysonics Randomic with 8 channels, Eigenmike with 32 channels, B&K with 36 channels, and VisiSonics with 64 channels.

For interactive sound reproduction applications, such as virtual reality, the listening positions might be changing from time to time. In this case, the way of using moving array or static array might be different. Under this circumstance, the moving array should be used to follow the main sound event [42]. The static array can also be used with advanced signal processing applied so that the reproduced sound adapts to sound event or user movements [29].

2.1.3. Binaural Recording

Binaural recording is an extended form of stereo recording. Ideally, binaural recording only captures the sound received at the left and right ear positions (eardrum for the dummy head and blocked or open entrance of the ear canal for human listeners). Therefore, binaural recording is the format that is the closest to the human hearing, when it is played through calibrated headphones. By recording sound at the eardrum positions, it automatically embeds all the cues needed for sound localization in the 3D space and the natural alteration of the sound timbre due to propagation, reflection, and scattering of sound as it interacts with our body.

The different binaural recording setups can be classified into four main types: (1) binaural microphones inserted in the ears of the human listeners; (2) microphones in the dummy head with torso, head and ears; (3) microphones in the simulator with head and ears; (4) microphones in the simulator with only ears.

First, a miniature microphone is usually placed at the entrance of the ear canal of human beings since it is impractical and difficult to be placed inside the ear canal near the eardrum. Examples of binaural microphones vary from professional, calibrated types (e.g., from Brüel & Kjær [43], shown in Figure 3) or low-cost types. Some other binaural recording microphones are integrated into a headset. An appropriate recording device is required to obtain the binaural recording.

In addition to the use of human beings for binaural recording, dummy heads are also commonly used in academia and the industry to obtain consistent and comparable results. A dummy head usually consists of a torso, head and ears, which are made up of special materials whose acoustic properties are similar to human body and its anthropometry generally follows closely with the average of the whole or a certain part of the population [44]. Several dummy heads that consist of a torso, head and ears include KEMAR (Knowles Electronics Manikin for Acoustic Research) [45], Brüel & Kjær 4128 HATS (head and torso simulator) [46], Head acoustics HMS IV [47], as shown in Figure 4. An example for dummy heads with only a head and two ears is the Neumann KU-100 [48]. As shown in Figure 5, the 3Dio binaural microphones [45] offer superb portability for binaural recording when compared to the bulky dummy heads, but at the expense of lacking head shadowing effects due to the absence of a head. Furthermore, the 3Dio omni binaural microphone records sound from four different head orientations, which could probably be more useful in interactive spatial audio in VR applications.



Figure 3. Brüel & Kjær 4101 Binaural microphone worn on the ear.



Figure 4. Dummy heads (from left to right): KEMAR, Brüel & Kjær 4128 HATS, Head Acoustics HMS III, and Neumann KU-100. Note that Neumann KU-100 dummy head is torso-free.



Figure 5. Binaural microphones from 3Dio, Free space binaural microphones.

2.1.4. Ambisonic Recording

Ambisonics is a method of recording and reproducing audio of a sound field in full-sphere surround. Essentially, ambisonics is a multi-channel surround sound format that does not contain any information or requirement for the playback configuration as demanded by other surround sound recording formats. This implies that the ambisonics recorded signals can be used in any playback setups. It is capable of full surround, including height and depth from a single point source in space [49].

Ambisonics format, or more precisely first order ambisonics (FOA), which is widely known as B-format surround signal, can be achieved by a special type of microphone in the shape of tetrahedron using four nearly coincident capsules. In other words, ambisonics can be simply understood as an extension of XY stereo recording with addition of the two other dimensions. The four components are labelled as W , X , Y and Z , where W is corresponding to an omnidirectional microphone and X , Y , Z corresponds to three spatial directions, i.e., front-back, left-right, and up-down, respectively, captured using figure-of-eight microphone capsules. For a given source signal, S with azimuth angle θ and elevation angle ϕ , ambisonics pans the desired four components as:

$$\begin{aligned} W &= \frac{S}{\sqrt{2}} \\ X &= S \cdot \cos \theta \cdot \cos \phi \\ Y &= S \cdot \sin \theta \cdot \cos \phi \\ Z &= S \cdot \sin \phi \end{aligned}$$

The resulting four-channel signal can then be transcoded into outputs of various formats, from a single source in mono to multichannel surround sound arrays. The major advantage being that with the initial capture, you can use post processing to vary the pan, tilt, zoom and rotation of the sound field, which is hard to achieve with other systems. The limitation of the first order ambisonics is the limited spatial resolution, which affects the sound localization and is only effective in a relatively smaller sweet spot. To improve the performance of first order ambisonics, higher order ambisonics are employed by adding more microphones to record sound field at higher orders. As a result, the reproduction of higher order ambisonics also requires more loudspeakers.

Examples of ambisonics microphones in the market include: Sennheiser AMBEO VR Microphone [50], Core Sound TetraMic [51], SoundField SPS200 Software Controlled Microphone [52], as shown in the Figure 6. Moreover, the spherical microphone arrays described in Section 2.1.2 can also be converted into first or higher order ambisonics format. For example, the Eigenmike consists of 32 microphones, which supports up to 4th order ambisonics. A comprehensive objective and subjective comparative study regarding the use of different ambisonics microphones was conducted in [53,54].



Figure 6. Ambisonics microphones (from left to right): Sennheiser AMBEO, Core Sound TetraMic, and SoundField SPS200.

2.2. Application of Spatial Audio Recording in Soundscape Studies

The section discusses how we can adopt different recordings techniques for soundscape studies. As the application of soundscape is relatively new, there are no comprehensive standards in terms of how recordings should be conducted for studies or reports on soundscape. Soundscape researchers employ different recording techniques for capturing the acoustic environment depending on various factors, such as the fidelity of reproduction medium, ease of capture, cost, etc.

Binaural and ambisonics are the two most common recording techniques in soundscape studies. Many have used binaural measurement systems, such as an artificial head or binaural microphones to record acoustic environment and binaural recordings are used for the reproduction of acoustic environments in laboratory-based listening experiments [54–56]. Ambisonic recording methods have recently received much attention in soundscape studies, as they are not restricted by playback mediums. Boren et al. [28] captured the acoustic environment in different parts of New York with an ambisonic microphone. Davies et al. [22] also used a FOA ambisonics microphone to record urban areas in Manchester, UK.

Although the mentioned recording methods for soundscape applications are not extensive, some insights can be gained to better understand how different recording techniques can be used in soundscape studies. Let us consider the three stages of the soundscape design process as introduced in Section 1. In Stage 1, recordings are used to evaluate the existing soundscape. In Stage 2, recordings are used to design a better soundscape. Finally, recordings are captured to validate the design soundscape after its implementation. In all these stages, the most critical requirement for soundscape recording is that it must sufficiently represent the characteristics of the acoustic environment in question that would facilitate the reproduction of the acoustic environment with sufficient perceptual accuracy. It should be made clear that the degree of perceptual accuracy is dependent on the goal of the study. On this note, the recording techniques must also be chosen with consideration of the reproduction mediums. A detailed review of the spatial audio reproduction techniques can be found in Section 3.

While the community continues to develop recording techniques for spatial audio reproduction, there are some trends and early examples that we can learn from. The strengths and weaknesses of these recording techniques in terms of their applications in soundscape studies are summarized in Table 1. With its simplicity, it is evident that ambisonics is the leading recording technique

for interactive spatial audio reproduction. On the other hand, conventional stereo and surround recording techniques can still be employed for specific applications, such as non-diegetic sound (or ambience) of the acoustic environment in soundscape studies. Microphone arrays are very useful to capture a more complete sound field (depending on the number of microphones used) and could be used in postprocessing to further emphasize certain sound components, such as, speech from noisy environment. Binaural recording still works great with static listening positions, but it does not allow interactions in soundscape studies. Even monophonic microphones are required for dry recordings that can be spatialized with post-processing, such as in auralization. Thus, it is clear that choosing the most suitable recording techniques depends on the reproduction techniques as well as the degree of perceptual accuracy required for soundscape studies. For high degree of perceptual accuracy, interactive spatial audio reproduction is required, and thus, ambisonic recordings might be a more suitable choice.

Table 1. Comparison of recording techniques for soundscape studies.

Recording Techniques	Strengths	Weaknesses	Remarks
Stereo and Surround recording	<ul style="list-style-type: none"> • Legacy recording methods; • Widely adopted in industry; • Direct playback over loudspeaker system. 	<ul style="list-style-type: none"> • Does not capture 3D sound field (usually only covers 1D or 2D); • Not suitable for direct playback over headphones; • Does not support head movements. 	<ul style="list-style-type: none"> • Limited spatial quality in soundscape studies.
Microphone arrays	<ul style="list-style-type: none"> • Ability to focus on certain sounds in different sound fields; • Support head movements. 	<ul style="list-style-type: none"> • Requires large number of microphones for good performance; • Requires sophisticated signal processing to obtain desired sound from recording. 	<ul style="list-style-type: none"> • A general method that could be used to record a more complete sound field for soundscape reproduction.
Ambisonics	<ul style="list-style-type: none"> • Records 3D sound fields with only 4 microphones; • Good mathematical foundations for recording and playback; • Efficient rendering for interactive applications; • Rapidly increasing popularity in industry. 	<ul style="list-style-type: none"> • Not suitable for non-diegetic sound like music; • Better performance requires higher order ambisonics; • Absence of international standards. 	<ul style="list-style-type: none"> • Well suited for interactive reproduction in soundscape studies.
Binaural recordings	<ul style="list-style-type: none"> • Closest to human hearing; • Direct playback over headphones. 	<ul style="list-style-type: none"> • Specialized equipment needed, e.g., ear simulators, or wearable binaural microphones; • Lack of support for head movements; • Non-personalized rendering (e.g., dummy head recordings). 	<ul style="list-style-type: none"> • Most commonly used recording technique for soundscape studies due to its simplicity; • Good spatial quality but limited interaction in soundscape studies • Personalized rendering (i.e., from in-ear binaural microphones).

3. Spatial Audio Reproduction Techniques for Soundscape Design

Perceptually accurate reproduction of the acoustic environment is crucial to achieve high ecological validity for evaluation of soundscape in laboratory conditions. This requires synthesis and rendering of different sound sources to create an immersive playback system, for instance, in subjective listening tests with sufficient perceptual accuracy. Soundscape composition and acoustic reproduction was pioneered by Schafer and his group in 1970s when they published a record titled “The Vancouver soundscape” [57]. As the reproduction techniques for spatial audio become more advanced, the soundscape researchers started adopting them in their studies from stereophonic techniques, multi-channel setups [57], to ambisonics and wave field synthesis [58].

A rendering algorithm along with multiple transducers are often used for reproduction of spatial sound. The sound can be rendered using either a pair of headphones or an array of loudspeakers arranged in a specific configuration. The aim of the system is to reproduce the sound field in such a

manner so as to give a sense of perception of spaciousness and directivity of sound objects located in 3D space [27]. The sounds generated in such a scenario are often referred to as virtual sound in contrast to real sound we commonly hear in everyday life [59].

The earliest attempt for reproduction of sound dates back to the phonograph invented in 1887. The first stereo system was introduced by Bluemin [60] in 1931 with its first commercial use in 1949. Dolby introduced the first surround sound system in 1970 and in the same year, Gerzon invented ambisonics [49]. In 1984, binaural reproduction technique is introduced by Sony and since 1989, the transaural technique gained popularity [61]. In 1993, another technique called the wave field synthesis (WFS) was proposed with its first major commercial application demonstrated in 2001 by Carrouso. In 2011, IOSONO came up with the world's first real-time spatial audio processor for WFS. In 2015, MPEG-H standard was announced, with the aim to ensure all types of sound formats are supported in any types of playback systems. The most recent attempts of spatial audio rendering were focused on VR/AR applications [27,62].

The above mentioned spatial audio reproduction techniques are primarily divided into two categories. The first category uses the technique of physical reconstruction of sound, which aims to synthesize the entire sound field in the listening area as close to the desired signal as possible. The second category is the perceptual reconstruction of sound, which employs psychoacoustic techniques to create a perception of spatial characteristics of sound [59]. The evolution of reproduction techniques is shown in Figure 7 along with their classification into physical and perceptual reproduction methods.

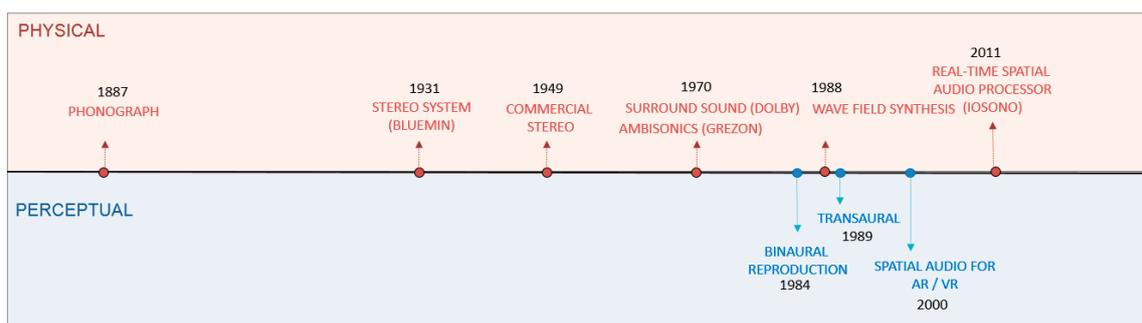


Figure 7. Evolution of spatial audio reproduction systems.

The primary aim of the section is to introduce the spatial audio techniques and to highlight the soundscape studies that have used these techniques. Moreover, since these techniques have different encoding and decoding formats, they require appropriate tools for rendering audio. Table A1 in Appendix A describes some of such tools that are available, where, many of these tools are open source and free to use. Usage of such tools in soundscape studies could enable the researchers to choose appropriate techniques according to the merits and demerits of each technique, which are described in the following Sections 3.1 and 3.2 and summarized in Table 2 in Section 3.3.

3.1. Physical Reconstruction of Sound

The aim of physical reconstruction of sound is to create the sound field in the listening area as close as possible to the desired sound field [59]. The oldest and one of the most popular methods of sound reproduction for soundscape uses two speakers in stereo configuration, i.e., placed at an angle of $\pm 30^\circ$ from the listener [60]. In soundscape studies, the stereo configuration was used by Schafer and his group. Vogel et al. [63] used stereophonic sound technique to study soundscape for French cities in 1997. Payne used a 2.1 channel system for testing the Perceived Restorativeness Soundscape Scale (PRSS) [64].

Multi-channel reproduction methods became popular in consumer devices from 1950. The international standard ITU-R BS.1116-1 establishes the recommendation for reproducing sound and assessment methods for multi-channel sound systems [65]. These methods have been widely used in the reproduction of acoustic environment in soundscapes. Gustavino et al. studied the subjective perception for playback through different multi-channel setups [66]. The speakers were arranged in different spatial configurations with a hexagon shaped prototype listening room created to test out different reproduction methods. The results of the experiments showed that while frontal perception was best for a 1-D configuration, the spatial definition of audio was best reproduced with 2-D configuration consisting of 6 speakers located at edges of the hexagonal room. In another article, Guastavino et al. [16] showed that multichannel reproduction is more suitable for processing complex auditory scenes, and playback of urban soundscapes in laboratory conditions, as compared to stereophonic reproduction techniques.

Wave field synthesis and ambisonics are the other two physical reconstruction techniques, which aim to create the same acoustical pressure field as present in the surrounding. Ambisonics, introduced by Gerzon et al. [49] is based on the decomposition of a sound field using spherical harmonics. Ambisonic reproduction is flexible to be applied in any sound playback configurations, which makes it very attractive to be used in a wide variety of applications [67]. Davis et al. [21] used the first order ambisonic microphones to record the background ambient soundscape and used a monophonic microphone to record the foreground sounds separately. The soundscape synthesis was carried out using a simulation tool. It allowed the sounds to be layered with each other and effects, like reverberation and reflections, to be added in real-time [68]. An eight-channel loudspeaker setup was used in a semi-anechoic chamber for playback. Boren et al. [58] described the usage of 16-channel audio configuration for playback of soundscape recording done in different parts of New York using sound field tetramic for recording. Moreover, there is increased support for ambisonics in VR/AR through leading industry players, including Google [69], BBC [70], and Facebook [71]. The ambisonics systems mentioned thus far are in the first-order configuration, with satisfactory but limited spatial resolution as compared to higher-order configurations, as discussed in Section 2.1.4. However, higher-order ambisonics systems are still pre-mature and costly today.

In a study for comparing the ecological validity among the different reproduction methods related to physical reconstruction of sound, Gustavino et al. [16] compared the stereophonic and ambisonic reproduction techniques. The study analyzed verbal data collected through questionnaires, and compared it to the field survey using semantic differential analyses for different sound samples. The study concluded that a “neutral visual environment” along with spatial immersion of recreated soundstage is essential to ensure high ecological validity.

3.2. Perceptual Reconstruction of Sound

Perceptual reconstruction techniques for spatial audio aim to replicate the natural listening experience by generating sufficient audio cues to represent the physical sound. Binaural technique, according to Blauert [72], is defined as “a body of methods that involve the acoustic input signals to both ears of the listener for achieving practical purposes, e.g., by recording, analyzing, synthesizing, processing, presenting and evaluating such signals”. Reproduction through this technique has two parts: one is the synthesis and rendering portion of the signal, and the other is the playback system. Head related transfer functions (HRTFs) are used to describe the change in sound spectrum due to the interaction of sound waves with listener’s body, head and pinna [59]. The synthesis and rendering of binaural audio is usually realized by convolving a dry sound source with the HRTF at a particular direction. For accurate reproduction, personalized binaural rendering using individualized HRTFs is required [73]. There are various techniques to accurately measure or synthesize the individualized HRTFs [74,75]. Recently, He et al. [76,77] proposed a fast and continuous HRTF acquisition system that incorporates the head-tracker to allow unconstrained head movements for human subjects. Binaural technique is used by soundscape researchers to playback the recorded sound with sufficient spatial

fidelity. Axelsson et al. [54] used it for their study with the aim of finding the underlying and primary components of soundscape perception. They used an artificial head (Brüel & Kjær Type 4100) for binaural recording and headphone (Senheiser HD600) for playback.

Binaural reproduction can also be realized over a pair of loudspeakers. Transaural audio is the method used to deliver correct binaural signals to ears of a listener using speakers (mostly in a stereo setup). However, there are some challenges to this approach, namely effective crosstalk cancellation and limited sweet spot range. Several solutions to these problems have been described in the literature, e.g., [78]. Often, double transaural approaches [78] are used to make sure that the front back confusion is minimum. However, this method has rarely been used in soundscape studies. Gustavino et al. [79] compared the quality of spatial audio reproduction between transaural, ambisonics and stereo methods. They used the Ircam default decoder to playback different auditory scenes from both indoors and outdoors [66], such as road traffic noise, a car interior, music concert, etc. Two experiments were performed, one for overall spatial quality evaluation and the other for localization accuracy. The results from the study indicate that ambisonics provide good immersion but poor localization for a sound scene. Stereophony and transaural techniques, on the other hand could be useful in the case where precise localization is required but do not have good immersive spatial sound.

3.3. Comparison of Different Spatial Audio Reproduction Techniques for Soundscape Studies

It is crucial that soundscape researchers select the appropriate sound reproduction technique for their use to make sure that the playback sounds similar to the natural listening scenario. To this end, Table 2 highlights the strengths and weaknesses of common reproduction techniques mentioned above, extended from studies in [80].

Table 2. Strengths and weaknesses of reproduction techniques for soundscape studies.

Reproduction Techniques	Number of Channels	Strengths	Weaknesses
Perceptual Reconstruction			
Binaural	Two	<ul style="list-style-type: none"> Enables creation of virtual source in 3D space with two channels. Lower equipment cost as compared to other solutions. 	<ul style="list-style-type: none"> It is only suitable for single user. Individualized HRTFs needed to avoid front back confusions and in-head localizations.
Transaural	Two or four	<ul style="list-style-type: none"> Enhances spaciousness and realism of audio with limited number of speakers. Accurate rendering of spatial images using fewer loudspeakers. 	<ul style="list-style-type: none"> Requires effective crosstalk cancellation. It is only suitable for single user.
Physical Reconstruction			
Stereo	Two	<ul style="list-style-type: none"> Legacy reproduction method. Widely adopted in industry. 	<ul style="list-style-type: none"> Poor spatial effect. The sound phantom image is always created at sweet spot.
Multichannel	Three or more	<ul style="list-style-type: none"> Better spaciousness of 360° audio as compared to stereo setups. Well adopted by industry. 	<ul style="list-style-type: none"> Large numbers of channels and speaker systems needed for spatial realism. Unable to achieve accurate accurate 360 degree phantom images.
Ambisonics	$(N + 1)^2$ for Nth Ambisonic	<ul style="list-style-type: none"> Can be used with any speaker arrangement. Core technology is not patented and free to use. 	<ul style="list-style-type: none"> Listener movement can cause artifacts in sound. Complicated setup. Not popular in industry.
Wave Field Synthesis	More than 100 usually	<ul style="list-style-type: none"> The sweet spot covers entire listening area. The virtual sound sources independent of listener position 	<ul style="list-style-type: none"> High frequency sounds beyond aliasing frequency not well reproduced. Large number of speakers needed.

In addition to the above techniques, Auralization techniques have recently become popular among soundscape researchers to achieve sufficient perceptual accuracy of acoustic reproduction. Auralization can be defined as a technique of creating audible sound files from numerical data [81]. It is not strictly a perceptual technique and can include the physical reproduction methods as well. The technique has been used in reproducing various acoustic environments, such as concert halls [82], auditoriums [83] and in room acoustics. The development of accurate models and efficient algorithms for accurate sound reproduction employs various techniques, like ray tracing [84], radiosity methods [85], finite element methods, etc. [86]. These models are used in conjugation with the reproduction setups to add audio effects and cues for increasing realism and spaciousness for spatial sound.

4. Spatial Audio in Virtual and Augmented Reality for Soundscape Design

Due to the importance of audio-visual interaction in soundscape perception, the audio techniques reviewed in this paper are commonly paired with a variety of visual mediums. Increasingly, virtual reality (VR) and augmented reality (AR) systems have been employed in soundscape research. Visuals recorded using omni-directional cameras, or designed in 3D modelling software, when rendered through head-mounted displays, are perceived to be immersive and realistic. To complement the immersive visuals, the accompanying audio has to possess a minimum degree-of spatialness to achieve natural realism. This argument stems from the perceptual nature of soundscapes and thus, a disjoint in the spatialness of the audio and visual elements will degrade the ecological validity [87]. Hence, the general rule of thumb for audio with VR and AR should correlate with the receivers' movements (mostly head movements), to create an interactive and immersive perception of soundscape.

Although spatial audio techniques for VR are discussed in detail in various studies [88,89], sufficient care is needed when applying these techniques into soundscape research. With the above rule in mind, the spatial audio recording techniques that is suitable for VR (or virtual sound in AR) must be able to capture the acoustic environment from all directions. From this perspective, ambisonics recording is the most commonly used recording techniques for VR/AR, though general microphone arrays with suitable post-processing can also be applied.

The spatial audio reproduction for VR/AR is usually realized through binaural rendering with head-tracking, though multichannel playback systems with an ambisonics decoder is also possible. In the binaural rendering system, when the human head moves, these HRTFs must be updated accordingly to account for the changes of the sound directions, which help to create a more interactive and natural listening experience. Note that the binaural rendering of the ambisonics recording is implemented by virtual loudspeakers in the ambisonics reproduction setup. For instance, a stationary VR scene that allows for head-track views of the visual scene (e.g., user is standing still in the virtual environment but able to turn the head to "look" around) needs to be accompanied by head-tracked binaural reproduction. An example of a reproduced VR scene with head-tracked binaural reproduction over headphones is shown in the right half of Figure 8. Recently, this approach has been used in soundscape research since it allows more immersive reproduction of the acoustic environment. One good example that soundscape research has benefitted from the use of VR techniques is in the SONORUS project [90]. Maffei et al. used VR for evaluating the influence of visual characteristics of barriers and wind turbine on noise perception [23,91,92].

Along with VR, the rising popularity of AR devices can be attributed to its availability in the consumer space. Devices such as the Hololens and Meta 2 have the capability to render holograms through head mounted displays [93,94]. With the consumerization of AR systems, soundscape researchers and practitioners have access to tools that can virtually augment visual and audio elements in the real environment. The general rule for applying audio to AR should still be adhered to, for achieving the desired perceptual accuracy during auralization in the soundscape design process.

There are several ongoing research works to achieve a high degree of perceptual accuracy for AR audio. Härmä et al. [95] described an augmented reality audio headset using binaural microphones to assist the listener with pseudo-acoustic scenes. Tikander et al. [96] also developed a mixer for equalizing

and mixing virtual objects with real environment. Ranjan et al. [27] proposed an augmented reality headset in which they use open-back headphones with pairs of external and internal microphones to achieve sound playback that is very similar to natural listening scenario.

These devices and upcoming innovations would be useful for projects that involve altering the soundscapes of existing locations. They could be used in soundscape studies to test different hypothesis in an immersive environment, which allows both virtual reproduced and real ambient sounds to be heard at the same time. AR devices have the advantage of including different elements of the perceptual construct of the soundscape, including meteorological conditions and other sensory factors [7]. Moreover, if accurate spatial sound is used with these AR devices, it would enable the soundscape researchers to fuse the virtual sound sources seamlessly with the real sound, thus enabling highly accurate interpretation of auditory sensation by the user. The viability of an augmented reality audio-visual system has been demonstrated with a consumer AR headgear and open-backed headphones with positional translation- and head-tracked spatial audio, as shown in the left half of Figure 8.



Figure 8. (Left) AR setup with spatial audio from rendered virtual sound objects for in situ environments. (Right) A lab-based VR setup from omni-directional camera recordings and a reproduced acoustic environment using spatial audio from ambisonic microphone recordings and rendered virtual sound objects [97].

In essence, VR systems with high fidelity spatial audio are well suited for an immersive acoustic environment with a good degree of control under laboratory conditions; whereas AR systems with high fidelity spatial audio are more immersive but are subject to a high degree of variability in the in situ environment. Hence, soundscape practitioners should exercise care in the selection of audio-visual mediums and should consider the complementary nature of VR and AR techniques mentioned.

5. Discussion

Soundscape design should be based on the relationship between human perceptual construct and physical phenomenon of the acoustic environment. As shown in Figure 1, acoustic recording and reproduction techniques are essentially adopted through every stage of soundscape design process. In particular, soundscape recording and reproduction techniques play a more critical role in Stage 2 for proposing and evaluating the soundscape designs before their implementations in situ. In Stage 2, various soundscape design approaches might be applied to improve the existing poor acoustic conditions, and those approaches should be assessed through subjective tests based on human perception of acoustic environment. In this context, soundscape design needs to consider the

degree of ecological validity of the soundscape for reliable solutions, which largely depends on the adopted recording and reproduction techniques. Therefore, for clarity, the discussion will be focused on highlighting the degree of perceptual accuracy of the recording and reproduction techniques introduced in Sections 2 and 3 respectively, in representing the acoustic environment.

As stated in ISO 12913-1:2014, any evaluation of soundscape is based upon the human perception of the acoustic environment (actual or simulated). The technique of constructing an audible acoustic environment, or Auralisation, is commonly used for soundscape evaluation in controlled environments. According to Vorländer [81], the goal of auralisation also stems from the human perception of the acoustic environment, in which the synthesized acoustics only needs to be perceptually correct and not physically perfect. Perceptual accuracy, also called plausibility, of the virtual acoustic environment is defined by Lindau and Weinzierl as “a simulation in agreement with the listener’s expectation towards an equivalent real acoustic event” [98]. The listener’s expectation, however, is subjective and can vary depending on the intended tasks in the acoustic environment. Ultimately, the ecological validity of a virtual acoustic environment will be application specific. This task-specific criterion implies that there are different levels of perceptual accuracy that can create “a suitable reproduction of all required quality features for a given application” [87]. Out of the quality features suggested by Pellegrini, emphasis will be directed to features, which are attributed to recording and reproduction techniques.

Therefore, it would be beneficial to classify the characteristics of the acoustic environment, such that the appropriate techniques are selected to achieve the desired level of perceptual accuracy. The characteristics of the acoustic environment are summarized in Table 3.

Table 3. Recommended audio reproduction and recording techniques for virtualizing/augmenting acoustic environments.

Characteristics of the Acoustic Environment					Recommended Techniques		
Spatial Fidelity ¹	Type of Environment ²	Movements		Virtual Sound Source Localization ³	Reproduction Techniques	Recording Techniques	Use Case(s) (Selected References, if Any)
		Listener Position ⁴	Head				
Low	Virtual (R/S)	×	×	0D	Mono loudspeaker; stereo headphone	Mono	Masking road traffic noise with birdsongs [99]
	Virtual (R/S)	×	×	1D	Stereo/surround loudspeaker; stereo headphone	Stereo/surround	Reproduced acoustic environment [25]; Perceived restorative-ness soundscape scale [71]
	Virtual (R/S)	×	×	2D	Surround sound loudspeakers with height Ambisonics (2D)	Array Ambisonics	Perception of reproduced soundscapes [22]
Med	Virtual (R/S)	×	×	3D–	Ambisonics; Binaural	Ambisonics; Binaural;	Auralising noise mitigation measures [100]; Masking noise with water sounds [101,102]
	Virtual (R/S)	×	×	3D+	Personalized binaural (PB) ⁵	Personalized binaural; Ambisonics ⁶	
	Virtual (R/S)	×	✓	3D+	Binaural/PB with head tracking	Ambisonics	
High	Virtual(S)	✓	✓	3D+	WFS; Binaural/PB with positional & head tracking	Mono (anechoic); Ambisonics	LISTEN project [103]
	Real + Virtual(S) ⁷	✓	✓	3D+	WFS; Binaural/PB with positional & head tracking	Mono (anechoic); Ambisonics	Augmented soundscape [27,97]

¹ Spatial fidelity refers to the quality of the perceived spatiality (location of sound in 3D space) of sound sources, which directly influences the auditory sensation. ² Acoustic environments can be real, captured, and reproduced (R), or synthesized from existing sound tracks (S). ³ As described in Section 5. ⁴ Refers to listeners’ translation movements. The virtual sound reproduction must be able to adapt the sound to translation movements. ⁵ Refers to binaural rendering with individualized HRTFs. ⁶ Requires convolution of down-mixed binaural with individualized HRTFs. ⁷ This refers to the case of recreating a virtual soundscape in a real acoustic environment. * Note that not all spatial audio recording and reproduction techniques reviewed have been used in soundscape studies, e.g., personalized binaural recording and rendering. The perceptual accuracy and ecological validity of these techniques need to be further examined.

The acoustic environment characteristics are organized firstly by its type, namely: (1) “Virtual(R)”, a simulated environment reproduced from recordings at a real location; (2) “Virtual(S)”, a simulated environment produced by auralization [81]; and (3) “Real”, the actual physical environment.

Next, the characteristics based on interactivity of the listener in the acoustic environment is identified. The interactivity or ‘Movement’ characteristics in the table are extremely relevant for VR and AR applications, as described in Section 4. For instance, the sound has to be in sync with the receivers’ positional and head movements as in the real environment (for AR) or in the virtual environment (in VR).

The ability to localize sound sources in the acoustic environment is also an important characteristic, which is also included and labelled as “Virtual sound source localization” in Table 3. The degree of localization is further categorized into 0D, 1D, 2D, 3D–, and 3D+. Virtual sound source localization in: (1) 0D refers to the perception with no spatial cues; (2) 1D refers to sound perception limited to left and right in horizontal plane; (3) 2D refers to sound perception with the inclusion of azimuthal and elevation dimensions; and (4) 3D refers to the sound perception with azimuth, elevation and distance, where – and + shows poor and good performance, respectively. Lastly, based on the degree of movement and localization characteristics, the spatial fidelity of the acoustic environment is simply classified into three grades for simplicity: Low, Medium, and High.

The description of the acoustic environment characteristics in Table 3, are accompanied by their respective reproduction and recording techniques. The limitations of the soundscape study, in terms of the spatial characteristics of the acoustic environment, can be decided by referring to Table 3. For reference, past soundscape evaluation studies are included in the last column of the table. It should be stressed that Table 3 does not suggest that all soundscape studies should employ high spatial fidelity audio, but instead provides a guide to construct the experimental conditions needed for soundscape research.

6. Conclusions

Recently, soundscape approaches have attracted more attention due to the increasing importance of evidence-based and perception-driven solutions to build better urban acoustic environments. Soundscape recording and reproducing techniques are essential tools for soundscape research and design. The present paper provides an overall picture of various spatial audio recording and reproduction techniques, which can be applied in soundscape studies and applications. Soundscape researchers should understand the strengths and weaknesses of these spatial audio recording and reproduction techniques, and apply the most appropriate techniques to suit their research purposes. Notably, the emerging VR/AR technologies, together with the advanced spatial audio recording and reproduction techniques, enable a more interactive and immersive auditory and visual perception, and would be a great fit for soundscape design with high ecological validity. Future research needs to focus more on spatial aspects of soundscape design elements for developing more accurate soundscape models.

Acknowledgments: Vectors modified under creative commons CC BY 2.0 for Figure 8 are designed by Vvstudio and Freepik. The background used in Figure 8 is modified for reuse under creative commons CC BY 3.0, was created by guillaumepaumier.com.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Tools for rendering spatial audio for soundscape studies.

Tools (Type of Application)	Reproduction Format/Rendering Usage							
	BIN	TRA	AMB	WFS	STE	VAR	MUL	AUR
Soundscape Renderer (C++ application for rendering) [104]	×		×	×				
Transpan (Max/MSP application) [105]	×	×						
Simmetry 3D (Multi channel reproduction) [106]							×	
PSTD (Blender interface for auralization) [107]								×
UrbanRemix (Collaborative soundscape measurement) [108]					×			
CATT acoustics (Room acoustics and auralization) [109]	×							×
EARS (Indoor auralization) [110]	×				×			×
Urban street auralizer (Micro scale urban areas simulation) [111]	×					×		×
FB360 spatializer (Sound spatializer for VR) [71]	×		×		×	×	×	
Google VR (Spatial sound plugin) [112]					×	×		×
Slab 3D plugin (Rendering plugin) [113]	×							×
Hololens spatial audio plugin (AR based spatial sound rendering) [94]						×		

BIN: Binaural, TRA: Transaural, AMB: Ambisonic first order WFS: Wave field synthesis, STE: Stereo, VAR: VR/AR, MUL: Multi channel surround Sound, VBAP: Vector Base Amplitude Panning, and AUR: Auralization.

References

- Pijanowski, B.C.; Farina, A.; Gage, S.H.; Dumyahn, S.L.; Krause, B.L. What is soundscape ecology? An introduction and overview of an emerging new science. *Landsc. Ecol.* **2011**, *26*, 1213–1232. [CrossRef]
- Guastavino, C. The ideal urban soundscape: Investigating the sound quality of French cities. *Acta Acust. United Acust.* **2006**, *92*, 945–951.
- Schulte-Fortkamp, B. Soundscapes and living spaces sociological and psychological aspects concerning acoustical environments. In Proceedings of the Forum Acusticum, Sevilla, Spain, 18 September 2002; p. 6.
- Hong, J.Y.; Jeon, J.Y. Influence of urban contexts on soundscape perceptions: A structural equation modeling approach. *Landsc. Urban Plan.* **2015**, *141*, 78–87. [CrossRef]
- Schafer, R.M. *The Soundscape: Our Sonic Environment and the Tuning of the World*; Destiny Books: New York, NY, USA, 1977.
- Kang, J.; Aletta, F.; Gjestland, T.T.; Brown, L.A.; Botteldooren, D.; Schulte-Fortkamp, B.; Lercher, P.; Van Kamp, I.; Genuit, K.; Fiebig, A.E.; et al. Ten questions on the soundscapes of the built environment. *Build. Environ.* **2016**, *108*, 284–294. [CrossRef]
- International Organization for Standardization. *Acoustics—Soundscape—Part 1: Definition and Conceptual Framework*; ISO 12913-1; ISO: Geneva, Switzerland, 2014.
- Herranz-Pascual, K.; Aspuru, I.; García, I. Proposed conceptual model of environmental experience as framework to study the soundscape. In Proceedings of the 39th International Congress and Exposition on Noise Control Engineering (INTERNOISE), Lisbon, Portugal, 13–16 June 2010; pp. 1–9.
- Zwicker, E.; Fastl, H. *Psychoacoustics: Facts and Models*; Springer Science & Business Media: Berlin, Germany, 2013; Volume 22.

10. Brown, A.L. A review of progress in soundscapes and an approach to soundscape planning. *Int. J. Acoust. Vib.* **2012**, *17*, 73–81. [[CrossRef](#)]
11. Aletta, F.; Kang, J.; Axelsson, Ö. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape Urban Plan.* **2016**, *149*, 65–74. [[CrossRef](#)]
12. Liu, F.; Kang, J. A grounded theory approach to the subjective understanding of urban soundscape in Sheffield. *Cities* **2016**, *50*, 28–39. [[CrossRef](#)]
13. Schulte-Fortkamp, B.; Fiebig, A. Soundscape analysis in a residential area: An evaluation of noise and people's mind. *Acta Acust. United Acust.* **2006**, *92*, 875–880.
14. Yang, W.; Kang, J. Soundscape and sound preferences in urban squares: A case study in sheffield. *J. Urban Des.* **2005**, *10*, 61–80. [[CrossRef](#)]
15. De Coensel, B.; Bockstael, A.; Dekoninck, L.; Botteldooren, D.; Schulte-Fortkamp, B.; Kang, J.; Nilsson, M.E. The soundscape approach for early stage urban planning: A case study. Proceedings of the Internoise, Lisbon, Portugal, 13–16 June 2010; pp. 1–10.
16. Guastavino, C.; Katz, B.F.G.; Polack, J.-D.; Levitin, D.J.; Dubois, D. Ecological Validity of soundscape reproduction. *Acta Acust. United Acust.* **2005**, *91*, 333–341.
17. International Organization for Standardization. *Acoustics—Soundscape Part 2: Data Collection and Reporting Requirements*; ISO 123913-2; ISO: Geneva, Switzerland, 2017.
18. Vorländer, M. *Auralization*; Springer-Verlag Berlin Heidelberg: Berlin, Germany, 2008.
19. Hong, J.Y.; Jeon, J.Y. Designing sound and visual components for enhancement of urban soundscapes. *J. Acoust. Soc. Am.* **2013**, *134*, 2026–2036. [[CrossRef](#)] [[PubMed](#)]
20. Kang, J.; Zhang, M. Semantic differential analysis of the soundscape in urban open public spaces. *Build. Environ.* **2010**, *45*, 150–157. [[CrossRef](#)]
21. Davies, W.J.; Bruce, N.S.; Murphy, J.E.; Bruce, N.S.; Murphy, J.E. Soundscape reproduction and synthesis. *Acta Acust. United Acust.* **2014**, *10*, 285–292. [[CrossRef](#)]
22. Sudarsono, A.S.; Lam, Y.W.; Davies, W.J. The effect of sound level on perception of reproduced soundscapes. *Appl. Acoust.* **2016**, *110*, 53–60. [[CrossRef](#)]
23. Ruotolo, F.; Maffei, L.; Di Gabriele, M.; Iachini, T.; Masullo, M.; Ruggiero, G.; Senese, V.P. Immersive virtual reality and environmental noise assessment: An innovative audio-visual approach. *Environ. Impact Assess. Rev.* **2013**, *41*, 10–20. [[CrossRef](#)]
24. Maffei, L.; Masullo, M.; Pascale, A.; Ruggiero, G.; Puyana Romero, V. On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces. *Energy Procedia* **2015**, *78*, 471–476.
25. Maffei, L.; Masullo, M.; Pascale, A.; Ruggiero, G.; Romero, V.P. Immersive virtual reality in community planning: Acoustic and visual congruence of simulated vs real world. *Sustain. Cities Soc.* **2016**, *27*, 338–345. [[CrossRef](#)]
26. Lacey, J. Site-specific soundscape design for the creation of sonic architectures and the emergent voices of buildings. *Buildings* **2014**, *4*, 1–24. [[CrossRef](#)]
27. Ranjan, R.; Gan, W. Natural listening over headphones in augmented reality using adaptive filtering techniques. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1988–2002. [[CrossRef](#)]
28. Boren, B.; Musick, M.; Grossman, J.; Roginska, A. I hear NY4D: Hybrid acoustic and augmented auditory display for urban soundscapes. In Proceedings of the 20th International Conference on Auditory Display, New York, NY, USA, 22–25 June 2014.
29. Eargle, J. *The Microphone Book: From Mono to Stereo to Surround—A Guide to Microphone Design and Application*; CRC Press: Boca Raton, FL, USA, 2012.
30. Bartlett, B.; Bartlett, J. *Practical Recording Techniques: The Step-by-Step Approach to Professional Audio Recording*; CRC Press: Boca Raton, FL, USA, 2016.
31. Huber, D.M.; Runstein, R.E. *Modern Recording Techniques*; CRC Press: Boca Raton, FL, USA, 2013.
32. Fletcher, H. Symposium on wire transmission of symphonic music and its reproduction in auditory perspective basic requirements. *Bell Syst. Tech. J.* **1934**, *13*, 239–244. [[CrossRef](#)]
33. Dunn, F.; Hartmann, W.M.; Campbell, D.M.; Fletcher, N.H.; Rossing, T.D. *Springer Handbook of Acoustics*; Rossing, T.D., Ed.; Springer: Berlin, Germany, 2015.
34. Benesty, J.; Chen, J.; Huang, Y. *Microphone Array Signal Processing*; Springer Science & Business Media: Berlin, Germany, 2008; Volume 1.
35. Rafaely, B. *Fundamentals of Spherical Array Processing*; Springer: Berlin, Germany, 2015; Volume 8.

36. Brüel & Kjær. Sound and Vibration Advanced Options for Holography and Beamforming. Available online: <http://www.bksv.com/Products/analysis-software/acoustics/noise-source-identification/spherical-beamforming-8606?tab=descriptions> (accessed on 5 December 2016).
37. Nokia OZO. Virtual Reality Camera with 360-Degree Audio and Video Capture. Available online: <https://ozo.nokia.com/> (accessed on 5 December 2016).
38. Algazi, V.R.; Duda, R.O.; Thompson, D.M. Motion-tracked binaural sound. *J. Audio Eng. Soc.* **2004**, *52*, 1142–1156.
39. mh Acoustics Eigenmike®Microphone. Available online: <https://mhacoustics.com/products> (accessed on 28 December 2016).
40. Visisonics Products Visisonics. Available online: <http://visisonics.com/products-2/#3daudio> (accessed on 3 February 2017).
41. Mathworks Phased Array Gallery—MATLAB & Simulink Example. Available online: <https://www.mathworks.com/help/phased/examples/phased-array-gallery.html> (accessed on 5 December 2016).
42. Paquette, D.; McCartney, A. Soundwalking and the Bodily Exploration of Places. *Can. J. Commun.* **2012**, *37*, 135–145. [CrossRef]
43. Brüel & Kjær. Sound & Vibration Type 4101-B-Binaural in-Ear Microphone Set. Available online: <http://www.bksv.com/Products/transducers/acoustic/binaural-headsets/4101?tab=overview> (accessed on 15 December 2016).
44. Vorländer, M. Past, present and future of dummy heads. In Proceedings of the Acustica, Guimarães, Portugal, 13–17 September 2004; pp. 1–6.
45. G.R.A.S. Head & Torso Simulators. Available online: <http://www.gras.dk/products/head-torso-simulators-kemar.html> (accessed on 13 June 2017).
46. Brüel & Kjær. Sound & Vibration TYPE 4128-C HATS. Available online: <https://www.bksv.com/en/products/transducers/ear-simulators/head-and-torso/hats-type-4128c> (accessed on 22 December 2016).
47. HEAD Acoustics Binaural Recording Systems—Artificial Head System HMS IV—Product Description. Available online: https://www.head-acoustics.de/eng/nvh_hms_IV.htm (accessed on 18 December 2016).
48. Georg Neumann GmbH Products/Current Microphones/KU 100/Description. Available online: https://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_description (accessed on 25 December 2016).
49. Gerzon, M.A. Width-height sound reproduction. *J. Audio Eng. Soc.* **1973**, *21*, 2–10.
50. Sennheiser. Sennheiser AMBEO VR MIC—Microphone 3D AUDIO Capture. Available online: <https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic> (accessed on 13 June 2017).
51. Core Sound TeraMic. Available online: <http://www.core-sound.com/TetraMic/1.php> (accessed on 28 December 2016).
52. SoundField. SoundField SPS200 Software Controlled Microphone | Microphones and Processors with Unique Surround Sound Capabilities. Available online: <http://www.soundfield.com/products/sps200> (accessed on 28 December 2016).
53. Bates, E.; Dooney, S.; Gorzel, M.; O'Dwyer, H.; Ferguson, L.; Boland, F.M. Comparing ambisonic microphones: Part 1. In *Audio Engineering Society Conference on Sound Field Control*; Audio Engineering Society: Guildford, UK, 2016.
54. Axelsson, Ö.; Nilsson, M.E.; Berglund, B. A principal components model of soundscape perception. *J. Acoust. Soc. Am.* **2010**, *128*, 2836–2846. [CrossRef] [PubMed]
55. Hong, J.Y.; Jeon, J.Y. The effects of audio-visual factors on perceptions of environmental noise barrier performance. *Landsc. Urban Plan.* **2014**, *125*, 28–37. [CrossRef]
56. Pheasant, R.; Horoshenkov, K.; Watts, G.; Barrett, B. The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places? *J. Acoust. Soc. Am.* **2008**, *123*, 1446–1457. [CrossRef] [PubMed]
57. Truax, B. Genres and techniques of soundscape composition as developed at Simon Fraser University. *Organ. Sound* **2002**, *7*, 5–14. [CrossRef]
58. Boren, B.; Andreopoulou, A.; Musick, M.; Mohanraj, H.; Roginska, A. I hear NY3D: Ambisonic capture and reproduction of an urban sound environment. In *135th Audio Engineering Society Convention (2013)*; Audio Engineering Society Convention: New York, NY, USA, 2004; p. 5.

59. He, J. *Spatial Audio Reproduction with Primary Ambient Extraction*; SpringerBriefs in Electrical and Computer Engineering; Springer Singapore: Singapore, 2017.
60. Blumlein, A.D. British Patent Specification 394,325. *J. Audio Eng. Soc. April* **1958**, *6*, 91–98.
61. Cooper, D.H.; Bauck, J.L. Prospects for transaural recording. *J. Audio Eng. Soc.* **1989**, *37*, 3–19.
62. Auria, D.D.; Mauro, D.D.; Calandra, D.M.; Cutugno, F. A 3D audio augmented reality system for a cultural heritage management and fruition. *J. Digit. Inf. Manag.* **2015**, *13*, 203–209.
63. Vogel, C.; Maffiolo, V.; Polack, J.-D.; Castellengo, M. Validation subjective de la prise de son en extérieur. In Proceedings of the Congrès Français D’acoustique 97, Marseille, France, 14–18 April 1997; pp. 307–310.
64. Payne, S.R. The production of a perceived restorativeness soundscape scale. *Appl. Acoust.* **2013**, *74*, 255–263. [[CrossRef](#)]
65. *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*; International Telecommunication Union: Geneva, Switzerland, 1997.
66. Guastavino, C.; Katz, B.F.G. Perceptual evaluation of multi-dimensional spatial audio reproduction. *J. Acoust. Soc. Am.* **2004**, *116*, 1105–1115. [[CrossRef](#)] [[PubMed](#)]
67. Fellgett, P. Ambisonics part one: General system description. *Media* **1975**, *17*, 20–22.
68. Bruce, N.S.; Davies, W.J.; Adams, M.D. Development of a soundscape simulator tool. In Proceedings of the 38th International Congress and Exposition on Noise Control Engineering 2009, INTER-NOISE, Ottawa, ON, Canada, 23–26 August 2009.
69. Omnitone. Available online: <http://googlechrome.github.io/omnitone/#home> (accessed on 13 June 2017).
70. BBC R & D Surround Sound with Height—BBC R & D. Available online: <http://www.bbc.co.uk/rd/projects/surround-sound-with-height> (accessed on 13 June 2017).
71. Facebook 360. Available online: <https://facebook360.fb.com/spatial-workstation/> (accessed on 12 June 2017).
72. Blauert, J. *An Introduction to Binaural Technology*; Gilkey, R.H., Anderson, T.R., Eds.; Springer: Berlin, Germany, 1997.
73. Ranjan, R. 3D Audio Reproduction: Natural Augmented Reality Headset and Next Generation Entertainment System Using Wave Field Synthesis. Ph.D. Thesis, School of Electrical and Electronic Engineering, Singapore, 2016.
74. Madole, D.; Begault, D. 3-D sound for virtual reality and multimedia. *Comput. Music J.* **1995**, *19*, 99. [[CrossRef](#)]
75. Sunder, K.; Tan, E.L.; Gan, W.S. Individualization of binaural synthesis using frontal projection headphones. *AES J. Audio Eng. Soc.* **2013**, *61*, 989–1000.
76. He, J.; Ranjan, R.; Gan, W.S. Fast continuous HRTF acquisition with unconstrained movements of human subjects. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 321–325.
77. Ranjan, R.; He, J.; Gan, W.-S. Fast continuous acquisition of HRTF for human subjects with unconstrained random head movements in azimuth and elevation. In Proceedings of the Audio Engineering Society Conference: 2016 AES International Conference on Headphone Technology, Aalborg, Denmark, 24 August 2016; pp. 1–8.
78. Gardner, W.G. *Transaural 3D Audio*; MIT Media Laboratory: Cambridge, MA, USA, 1995.
79. Guastavino, C.; Larcher, V.; Catusseau, G.; Boussard, P. Spatial audio quality evaluation: Comparing transaural, ambisonics and stereo. In Proceedings of the 13th International Conference on Auditory Display, Montréal, QC, Canada, 26–29 June 2007; pp. 53–58.
80. SIGGRAPH. “Sounds Good to Me” SIGGRAPH 2002 Course Notes. Available online: <https://www.cs.princeton.edu/~funk/course02.pdf> (accessed on 12 June 2017).
81. Vorländer, M. *Auralization: Fundamentals of Acoustics, Modelling, Simulation*; Springer: Berlin, Germany, 2008.
82. Pätynen, J.; Lokki, T. Evaluation of concert hall auralization with virtual symphony orchestra. In Proceedings of the International Symposium on Room Acoustics, Melbourne, Australia, 29–31 August 2010; pp. 1–9.
83. Rindel, J.H. Modelling in auditorium acoustics—from ripple tank and scale models to computer simulations. In Proceedings of the Forum Acusticum, Sevilla, Spain, 19 September 2002; pp. 1–8.
84. De Coensel, B.; de Muer, T.; Yperman, I.; Botteldooren, D. The influence of traffic flow dynamics on urban soundscapes. *Appl. Acoust.* **2005**, *66*, 175–194. [[CrossRef](#)]

85. Kang, J. Numerical modeling of the sound fields in urban squares. *J. Acoust. Soc. Am.* **2005**, *117*, 3695–3706. [[CrossRef](#)] [[PubMed](#)]
86. Hothersall, D.C.; Horoshenkov, K.V.; Mercy, S.E. Numerical modelling of the sound field near a tall building with balconies near a road. *J. Sound Vib.* **1996**, *198*, 507–515. [[CrossRef](#)]
87. Pellegrini, R.S. Quality assessment of auditory virtual environments. In Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, 29 July–1 August 2001; pp. 161–168.
88. Begault, D.R. *3-D Sound for Virtual Reality and Multimedia*; Academic Press Professional, Inc.: San Diego, CA, USA, 2000.
89. Sunder, K.; He, J.; Tan, E.L.; Gan, W.S. Natural sound rendering for headphones: Integration of signal processing techniques. *IEEE Signal Process. Mag.* **2015**, *32*, 100–113. [[CrossRef](#)]
90. Alves, S.; Estévez-Mauriz, L.; Aletta, F.; Echevarria-Sanchez, G.M.; Puyana Romero, V. Towards the integration of urban sound planning in urban development processes: The study of four test sites within the SONORUS project. *Noise Mapp.* **2015**, *2*, 57–85.
91. Maffei, L.; Masullo, M.; Aletta, F.; Di Gabriele, M. The influence of visual characteristics of barriers on railway noise perception. *Sci. Total Environ.* **2013**, *445–446*, 41–47. [[CrossRef](#)] [[PubMed](#)]
92. Maffei, L.; Iachini, T.; Masullo, M.; Aletta, F.; Sorrentino, F.; Senese, V.P.; Ruotolo, F. The effects of vision-related aspects on noise perception of wind turbines in quiet areas. *Int. J. Environ. Res. Public Health* **2013**, *10*, 1681–1697. [[CrossRef](#)] [[PubMed](#)]
93. Meta 2. Available online: <https://www.metavision.com/> (accessed on 22 March 2017).
94. Microsoft. Microsoft HoloLens. Available online: <https://www.microsoft.com/microsoft-hololens/en-us> (accessed on 12 June 2017).
95. Härmä, J.; Jakka, M.; Tikander, M.; Karjalainen; Lokki, T.; Hiipakka, J.; Gaëtan, L. Augmented reality audio for mobile and wearable appliances. *J. Audio Eng. Soc.* **2004**, *52*, 618–639.
96. Tikander, M.; Karjalainen, M.; Riikonen, V. An augmented reality audio headset. In Proceedings of the International Conference on Digital Audio Effects (DAFx-08), Espoo, Finland, 1–4 September 2008; pp. 181–184.
97. Gupta, R.; Lam, B.; Hong, J.; Ong, Z.; Gan, W. 3D audio VR capture and reproduction setup for auralization of soundscapes. In Proceedings of the 24th International Congress on Sound and Vibration, London, UK, 23–27 July 2017.
98. Lindau, A.; Weinzierl, S. Assessing the plausibility of virtual acoustic environments. *Acta Acust. United Acust.* **2012**, *98*, 804–810. [[CrossRef](#)]
99. Hao, Y.; Kang, J.; Wörtche, H. Assessment of the masking effects of birdsong on the road traffic noise environment. *J. Acoust. Soc. Am.* **2016**, *140*, 978–987. [[CrossRef](#)] [[PubMed](#)]
100. Thomas, P.; Wei, W.; Van Renterghem, T.; Botteldooren, D. Measurement-based auralization methodology for the assessment of noise mitigation measures. *J. Sound Vib.* **2016**, *379*, 232–244. [[CrossRef](#)]
101. Galbrun, L.; Ali, T.T. Acoustical and perceptual assessment of water sounds and their use over road traffic noise. *J. Acoust. Soc. Am.* **2013**, *133*, 227–237. [[CrossRef](#)] [[PubMed](#)]
102. Jeon, J.Y.; Lee, P.J.; You, J.; Kang, J. Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. *J. Acoust. Soc. Am.* **2010**, *127*, 1357–1366. [[CrossRef](#)] [[PubMed](#)]
103. Eckel, G. Immersive audio-augmented environments the LISTEN project. In Proceedings of the Fifth International Conference on Information Visualisation, London, UK, 25–27 July 2001; pp. 571–573.
104. Geier, M.; Ahrens, J.; Spors, S. The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. In Proceedings of the Audio Engineering Society—124th Audio Engineering Society Convention, Amsterdam, The Netherlands, 17 May 2008; pp. 179–184.
105. Baskind, A.; Carpentier, T.; Noisternig, M.; Warusfel, O.; Lyzwa, J. Binaural and transaural spatialization techniques in multichannel 5.1 production. In Proceedings of the 27th Tonmeistertagung, VDT International Convention, Cologne, Germany, 22–25 November 2012.
106. Contin, A.; Paolini, P.; Salerno, R. *Sensory Aspects of Simulation and Representation in Landscape and Environmental Planning: A Soundscape Perspective*; Springer International Publishing AG: Cham, Switzerland, 2008; Volume 10, pp. 93–106.
107. OpenPSTD Homepage. Available online: <http://www.openpstd.org/> (accessed on 22 March 2017).
108. Freeman, J.; DiSalvo, C.; Nitsche, M.; Garrett, S. Soundscape composition and field recording as a platform for collaborative creativity. *Organ. Sound* **2011**, *16*, 272–281. [[CrossRef](#)]

109. CATT Acoustic. CATT Acoustic Software v9.1. Available online: <http://www.catt.se/CATT-Acoustic.htm> (accessed on 22 March 2017).
110. Ahnert, W.; Feistel, R. EARS Auralization Software. *J. Audio Eng. Soc.* **1993**, *41*, 894–904.
111. Kang, J.; Meng, Y.; Brown, G.J. Sound propagation in micro-scale urban areas: Simulation and animation. *Acta Acust.* **2003**, *89*, 1–6. [[CrossRef](#)]
112. Google. Google VR Spatial Audio Homepage. Available online: <https://developers.google.com/vr/spatial-audio> (accessed on 22 March 2017).
113. Slab3D Software Homepage. Available online: <http://slab3d.sourceforge.net/> (accessed on 22 March 2017).



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).