KRISTINA GOLUBKOVA A REVIEW OF SPATIAL AUDITORY SCENE ASSESSMENT Bilkent University 2022

A REVIEW OF SPATIAL AUDITORY SCENE ASSESSMENT APPROACHES

A Master's Thesis

by
KRISTINA GOLUBKOVA

Department of
Music
İhsan Doğramacı Bilkent University Ankara
August 2022

A REVIEW OF SPATIAL AUDITORY SCENE ASSESSMENT APPROACHES

The Graduate School of Economics and Social Sciences of
İhsan Doğramacı Bilkent University

by

KRISTINA GOLUBKOVA

In Partial Fulfillment of the Requirements for the Degree of
MASTER OF ARTS IN MUSIC

THE DEPARTMENT OF
MUSIC
İHSAN DOĞRAMACI BİLKENT UNIVERSITY
ANKARA
August 2022

A REVIEW OF SPATIAL AUDITORY SCENE ASSESSMENT APPROACHES
By Kristina Golubkova

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Music.

Onur Türkmen
Advisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Music.

Onur Yıldırım
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Music.

Tolga Tüzün
Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences

----

Refet Soykan Gürkaynak
Director

ABSTRACT

A REVIEW OF SPATIAL AUDITORY SCENE ASSESSMENT APPROACHES

Golubkova, Kristina
M.A., Department of Music
Advisor: Associate Prof. Dr. Onur Türkmen

August 2022

The state-of-the-art research in sound technologies is now increasingly focused on augmented reality systems, which use spatial auditory content to represent information beyond that encountered in the real world. Successful implementation of spatial sound on such a large scale requires comprehensive and reliable research data in acoustics, cognition, and the principle of auditory processing. The main contribution of Rumsey's scene-based paradigm is a theoretical framework for the description and analysis of space-related scenes.

This thesis presents an overview of the existing research related to the perception of spatial sound within three domains: spatial hearing and cognition, spatialisation, and evaluation of spatial quality attributes.

Keywords: Spatial Perception, Auditory Quality Assessment, Auditory Scene Analysis, Spatial Audio Reproduction

ÖZET

MEKAN BAZLI İŞITSEL SAHNELERIN ANALIZI ÜZERINE YAKLAŞIMLARIN BIR DEĞERLENDIRMESI

Golubkova, Kristina
Yüksek Lisans, Müzik
Tez Danışmanı: Doç. Dr. Onur Türkmen

Ağustos 2022

Ses teknolojileri üzerine yapılan güncel araştırmalar yükselen bir oranda artırılmış gerçeklik sistemleri üzerine yoğunlaşıyor. Bu sistemler, gerçek yaşamda karşılaşılanın ötesinde mekan bazlı işitsel veriler kullanıyor. Mekansal sesin bu denli büyük bir ölçekte başarılı uygulanabilmesi için akustik, bilişsel algı, ve işitsel işlemleme presnipleri üzerine kapsayıcı ve güvenilir araştırma verileri gerekmekte. Rumsey'in sahne tabanlı paradigması ise mekan bazlı sahnelerin tasvir ve analizi için teorik bir çerçeve sunmakta.

Bu tez, mekansal ses algısı üzerine halihazırda var olan araştırmaları üç alan içerisinde inceliyor: mekansal işitim ve bilişsellik, mekansallaştırma, ve mekansal özelliklerin değerlendirilişi.

Mekansal Algi, Isitsel Özelliklerin Degerlendirmesi, Isitsel Sahne Analizi, Mekansal Ses Üretimi

ACKNOWLEDGEMENTS

I wish to thank Onur Türkmen for his academic supervision. I would like to thank Tolga Tüzün for academic support. And many thanks to my family and friends for their moral support.

TABLE OF CONTENTS

INTRODUCTION

Space is researched in various sound-related fields of science and application, including acoustics, architectural and environmental noise control, auditory physiology and perception, music cognition and performance, psychoacoustics and hearing aid design. One of the main vectors is spatial sound research is obtaining solid data of the human auditive perception with the intention to implement the models in reproduced and virtual sonic environments.

The main focus of spatial sound research is on how humans perceive the direction and distance of a sound source, the subjective evaluation of audio quality, and how perception can be effectively manipulated with technological means.

The aim of the thesis is to provide an outlook on the existing body of research regarding perception of space within three domains: spatial hearing and cognition, spatialisation techniques, and spatial attribute evaluation in reproduced audio.

The thesis is divided into three chapters:

The first chapter provides a general introduction to the perception of space and provides an overview of basic acoustic and psychoacoustic principles. It mainly focuses on the perception of space from a cognitive perspective, including the mechanisms behind auditory scene analysis and auditory streaming.

The second chapter discusses different fundamental principles of spatialisation techniques and their effect on perceived spatiality. It focuses on the use of reverberation, binaural cues, head-related transfer functions and acoustic diffuseness in reproduction systems.

CHAPTER I

AUDITORY PERCEPTION AND COGNITION

1.1 Auditory Spatial Perception

Auditory spatial perception is the capacity to recognize particular sound sources in three-dimensional space, especially in situations where multiple sound sources occur simultaneously (R. Roginska & Geluso, 2018, pp. 10–11).

The study of spatial hearing revolves around the research of two categories of characteristics:

1) of the source signal, and
2) that of the acoustic environment (Bates, 2009, p. 14).

Both of these categories and the perceptual mechanisms behind them are important to consider when attempting to assess any spatial sound scene in order to create a plausible spatial illusion. Therefore, one of the main mechanisms of spatial hearing is the ability to localise the sound source, which involves perceived direction and the distance to it, as well as the reflections created by the signal diffusion in the environment (Bates, 2009, p. 14). In the preliminary stage of auditory processing, a complex acoustic signal is modified and distorted by complex acoustic modifications and distortion of it by the pinna, ear canal, and middle ear (Warren, 2008, p. 35). The auditory periphery alone does not have a mechanism for eliciting dimensional information from the perceived sound; it is the function of the neural paths to further compute the perceived localisation (Yost, Popper, & Fay, 2008, p. 7).

The physiology of the auditory system and cognitive processes are both responsible for the multi-faceted phenomena of auditory perception. When combined with information from the other somatosensory systems, neural activity transforms independent aspects of auditory input, such as spectral content, temporal properties, and location, into distinct perceptual representations. (Roginska & Geluso, 2018).

"Localisation is the process of linking spaces, namely that of linking the position of a given physical source, with that of the "position" of the listener's auditory event (if any)" (Suzuki et al., 2011, p. 3.) Commonly, the arriving auditory information affecting localisation mapping is called *localisation cues*. Gilkey and Anderson (1997) define a potential localisation cue as "any physical aspect of the acoustical waveforms reaching a listener's ears that is altered by changes in the position of the sound source relative to that of the listener" (Gilkey & Anderson, 1997, p. 2). The term localisation is particularly applied to describe the location of externalised sources in space (Warren, 2008, p. 35).

When describing the position of a sound source, it is common to use terms such as "azimuth (angle from straight ahead measured in the horizontal plane), elevation (angle from the horizontal measured in a vertical plane), and distance" (Warren, 2008, p. 35). These three degrees of spatial information correspond to the three dimensions in which sound can occur: left-right, front-back, and up-down.

In order to determine spatial location, the auditory system uses *binaural cues* and *frequency-dependent pinna filtering* (R. Roginska & Geluso, 2018, pp. 10–11). Further, particular localisation cues will be reviewed.

1.1.1 Localisation Cues in the Horizontal Plane

Fundamental to directional hearing is the left-right axis. Left-right positioning is the least susceptible to inaccuracy, and the highest directional acuity is centred along the horizontal plane (Kendall, 2010, p. 230). The accuracy of such positioning is determined by the mechanisms of interaural differences.

Interaural spectrum discrepancies are caused by wave diffraction and reflection between the ears. Since the ears are positioned laterally, interaural differences cannot be utilised to identify front/back orientation without head movement. We use minute spectral changes caused by forward-pointing pinnae to distinguish the front from rear sources (Khosravi, 2012, p. 28).

Lord Rayleigh's "duplex theory" (Roginska & Geluso, 2018, p. 11) emphasises the relevance of two fundamental signals—interaural time differences (ITDs) and interaural intensity differences (IIDs). Independent of frequency content, variations in the overall difference between left and right intensity levels at the eardrum are interpreted as changes in the location of the sound source from the perspective of the listener (Roginska & Geluso, 2018, p. 16).

Another dimensional axis on the horizontal plane is the front-back azimuth. Unlike the left-right location mechanism, the primary localisation cues for front-back position detection are monaural, i.e. the pinnae alter the intensity in particular frequency regions depending on the signal's position within the median plane (Blauert, 2013, p. 18). The difference between front and rear is highly reliant on the dynamic head movement. The front–back difference is secondary to the acoustics of the forward-facing pinnae, which cause spectral distinctions between front and rear sound events (Kendall, 2010, p. 230).

1.1.2 Sound Localisation Cues in the Vertical Plane

The exact mechanism of monaural filtering applies to elevation detection. The interaural time difference cannot act as an elevation indicator since the "pinna is not symmetrical along its vertical axis" (Khosravi, 2012, p. 28). This asymmetrical acoustics gives elevation indications, while directional hearing above the head is inaccurate and nearly nonexistent below (Kendall, 2010, p. 231).

1.1.3 Fundamental Cues to Distance

Sound level and reverberation determine a sound source's perceived distance. Sound sources reach a listener through *direct sound*, reflection, and diffraction off the environment. As a multi-dimensional percept, perceived distance requires processing of various cues, such as loudness, spectral content, reverberation, and cognitive familiarity (Roginska & Geluso, 2018, pp. 10–11).

As the source and listener distance increases, direct and early reflected sound levels change. However, diffuse reverberation is essentially independent of listener position. Beyond a certain distance, the reverberant signal level is higher than direct signal level, and perceived distance becomes constant regardless of the source distance. This proves that the auditory system estimates the direct-to-reverberant ratio using directional binaural cues to differentiate direct and reverberant sounds. While source signal amplitude and the relationship between direct and indirect signals are prominent cues in distance perception, various secondary cues also indicate relative distance. Due to air absorption, increasing source distance reduces the signal's high-frequency spectrum richness (Bates, 2009, p. 23).

The ratio of direct-to-indirect, or reflected, sound in reverberant environments changes with the listener's distance from the sound source. With increasing distance, the ratio of indirect sound in the total sound image increases. At a certain distance, the so-called "reverberation distance", the direct and reflected sound levels are equal. At or beyond the critical distance or beyond, the total sound intensity tends to remain constant due to the dominance of indirect sound (Roginska & Geluso, 2018). Although the direct sound is followed by a number of isolated reflections that can be heard, the first wavefront dominates many aspects of perception. A group of processes known as the "precedence effect" are thought to be involved in resolving perception and localization issues between a direct sound and a reflection (Litovsky, Colburn, Yost, & Guzman, 1999).

The amplitude of a sound object is considered to be the most prominent cue to the distance to it. However, the degree of familiarity with the source signal has an impact on the salience of the perceived distance (as cited in Bates, 2009, p. 22), with unfamiliar sounds providing less stable cues for the detection of distance.
When anticipating or becoming accustomed with the sound source, particularly speech, distance cues may fluctuate. Even without spatial clues, one may detect the distance to distinct sounds and their context (Roginska & Geluso, 2018, p. 25).

1.1.4 Environmental Context Cues

Environmental context permits us to differentiate between the same source in different enclosure sizes or in the open air. The spatial distribution of early reflections from environmental sounds is equally as essential as reverberation time and dire-to-indirect sound ratio (Bronkhorst & Houtgast, 1999; Kendall & Martens, 1984). These characteristics can assist in defining the context and distance range of a sound source. Environmental background noise can aid in determining the range and context of sounds. According to research, distance and azimuth calculations improve as listeners "learn" the environmental situation (as cited in Roginska & Geluso, 2018, p. 30).

It can be concluded that a multitude of factors may alter the perceived location of a sound source in auditory space, namely the frequency content of the sound signal, familiarity with the sound sources, and the nature of the reflective environment (Gilkey & Anderson, 1997, p. 1).

1.2 Perceptual Cues in a Complex Auditory Environment

In a complex auditory environment, multiple sound signals arrive at the ear concurrently. The so-called "cocktail party effect", introduced by Cherry in his article on selective hearing (Cherry, 1953, p. 976), describes a mechanism whereby the listener can selectively focus on one of the simultaneous, overlapping streams of conversation. This research was a precursor to the idea of *auditory scenes* and methods of grouping and segregation of complex auditory events, developed and thoroughly researched by Bregman (A. Bregman, 1994).

Auditory scene analysis is a subfield of psychoacoustic studies that focuses on assessing auditory events and objects contained in a scene by eliciting particular sonic components in human cognitive processes and their relations with other objects and the environment that contains them.

The formation of auditory concepts appears to be linked to perceptual streaming. Specific time and frequency relationships between signal components prompt the brain to associate them as a single object. When an object has been previously identified and labelled, memory allows for its identification. Streaming is a component of higher-level pattern recognition in perception (Rumsey & Safari, 2012, p. 43). Such principles of object grouping are addressed in the context of visual perception in the field of gestalt psychology (Deutsch, 2013). The principles proposed by the early Gestalt psychologists included those of proximity, similarity, good continuation, and common fate (Wagemans et al., 2012).

Bregman describes several principles upon which incoming complex sounds can be grouped into so-called *streams* (A. S. Bregman & McAdams, 1994).

1.2.1 The Principle of Proximity and Similarity

The grouping principle of proximity and similarity refers to the physical parameters of the sounds, such as frequency, temporal onset, intensity, location, and spectral shape. Sounds that are close and/or similar with regard to these criteria tend to be grouped together (*stream fusion*) and will be perceived as separate streams otherwise (*stream segregation*) (Gelfand, 2018, p. 324).

1.2.2 The Principle of Common Fate and Good Continuation

Common fate refers to temporal changes of the stimuli that include micro-modulation, vibrato, and synchronised variation in frequency. Good continuation presupposes an uninterrupted, smooth counter of changes in the parameters. Stimuli are perceived to have a common origin if these two parameters can be applied (A. Bregman, 1994).

1.2.3 Primitive and Schema-based Segregation

Additionally, Bregman describes the difference between two modes of auditory streaming, the primitive and schema-based approaches (A. Bregman, 1994). Primitive segregation operates in a bottom-up fashion, where the incoming signal is split into multiple streams according to temporal and spectral criteria and consequently regrouped into events with a perceptually common origin. With this, grouping according to temporal criteria is called sequential integration, which includes criteria of temporal proximity, good continuation, and spectral shape, while according to spectral is referred to as simultaneous integration and involves principles of common fate and perceived spatial origin (A. S. Bregman, Ahad, Crum, & O'Reilly, 2000). The primitive segregation is described by Bregman as innate, automatic, and therefore symmetrical — as the process does not focus on a particular aspect of the incoming signal.

In contrast, schema-based segregation is a top-down approach, operating on the pre-learned concepts, consciously involving past experiences, attention, and expectations to operate on complex auditory streams. Therefore, the schema-based process is considered to be asymmetrical and voluntary, which, however, results in longer time of processing compared to the primitive one (Gelfand, 2018, p. 325).

The top-down approach is also said to operate across modalities, such as visual reference for sound localisation (Marian, Hayakawa, & Schroeder, 2021, p. 1) that employs information from different sensory modalities to construct a "*unitary event*" (Vroomen & Gelder, 2000, p. 2).

1.2.4 Spatiality of Events

In Bregman's research on auditory scenes, spatial cues are given secondary significance in determination of grouping patterns. He writes:

> "In audition it is frequency for which the representation is laid out like a map in the peripheral levels of the system. At higher levels of the nervous system, though, there are spatial maps of sound. Evidence has been found that at the level of the superior colliculus of the nervous system the spatial origin of

auditory inputs is coded topographically" (A. Bregman, 1994, p. 73).

This ability to localise sounds and assess spatial auditory scene is not exclusively confined to the processes of spatial hearing but rather "appear to arise from a deeply embodied knowledge of space and spatial movement" (Kendall, 2010, p. 228). Cognition does not operate independently on stimuli received from different sensory modalities. Instead, "it is conceived of as an active process in which goal-directed actions and sensory feedback constitute a continuous dynamic pattern of sensorimotor activation" (Küssner, 2014, p. 21). The co-influence of different sensory modalities plays a crucial role in processing of complex physical and mental tasks.

Spatial perception of sounds mainly relies on information from sensory-motor and vision modalities. "The spatial hearing facilitation that can result from seeing the overall environment, may reflect the active sensory-motor exploration of the auditory scene" (as cited in Marian et al., 2021, p. 2). In this regard, Kendall (Kendall, 2010, p. 229) describes sound events as "intrinsically spatial", as they occur in space. The embodied concepts of gravitations, inertia, and motion patterns are inseparably integrated into the process of complex auditory scene assessment (Boren & Musick, 2018, p. 36). They are extensively employed in analyses of electroacoustic music to conceptually describe various spectromorphologies (Calderón-Garrido & Gustems, 2015; Smalley, 1997).

The section describes the main principles of human sound localisation and auditory stream segregation. It is clear that the processing of complex auditory scenes employs physiological, neural, and cognitive processes, innate and conscious, in order to construct a mental percept of the sound events in space. The auditory system is thought to process sounds in a hierarchical fashion, starting with low-level features such as amplitude and frequency, then proceeding to higher-level features, including spectral shape and temporal envelope.

CHAPTER II

SPATIALISATION METHODS AND IDIOSYNCRASIES

In the context of sound reproduction, the term *spatial audio* refers to the concept of sound being in a three-dimensional space surrounding a listener. The primary focus of spatial audio is on the location-specific characteristics of a sound or sound field. While the term *sound localisation* concerns how listeners perceive location-specific characteristics of the sound field in the environment, the term *sound spatialisation* relates to how sound is dispersed throughout a particular reproduction system (Carty, 2009, p. 282). Eventually, spatialisation aims to create a convincing spatial illusion by synthesising “spaces and spatial properties of sounds” (Peters, Marentakis, & McAdams, 2011, p. 10).

Spatialisation techniques utilise parameters of spatial dimensions, location of the sound sources and their direction, along with the acoustic properties of the physical environment where the sound work is meant to be performed. In this sense, space and spatialisation are mutually dependent, where physical space dictates the choice of a particular spatialisation approach, through which the space is given desired function and character (Harley, 2016, p. 4). Thus, particular attention should be paid to the features and idiosyncrasies of multichannel sound reproduction that contribute to the production of spatial sound images.

The primary component involved in sound localisation is binaural hearing. Interaural Time Difference provides the brain with two independent signals, which essentially contribute to sound localisation in horizontal plane. Monaural cues, in turn, provide critical information for elevation and front-back localisation. Another significant phenomenon in spatial sound reproduction, particularly with multi-speaker setups, is the precedence effect; that is, in the situation where multiple similar signals are heard, the auditory systems prioritise the first arriving wavefront to localise the sound event (Carty, 2009, p. 282).

The invention of sound recording in the 19th century initiated the development of audio reproduction technologies, which progressively increased the number of channels, configurations, and later synthesis techniques, thus adding to the perceived dimension of the sound (Roginska & Geluso, 2018, p. 64). Monophonic reproduction is essentially not capable of conveying a convincing spatial impression. The tools for creating some depth and distance were limited to reverberation (Rumsey & Safari, 2012, p. 10). In the 1930s, the initial concepts for recreating missing spatial information were recognised; however, they were still limited by technological and financial constraints. In 1934, The Bell Telephone Laboratories proposed two novel approaches that could reproduce spatial dimensions, namely binaural and multichannel techniques (Toole, 1997, p. 2).

2.1 Multichannel Reproduction Techniques

In multichannel systems, each channel is associated with a loudspeaker which creates easily localisable sound sources and is commonly referred to as a stereophonic setup (Toole, 1997, p. 2).

The illusion of spatial sound image in stereo systems is achieved by distributing directional "sound sources between two or more loudspeakers" (Roginska & Geluso, 2018, p. 64) to the listener at the ideal listening position, i.e. the sweet spot. The listener in the sweet spot is centred between and facing the loudspeakers equidistantly to form a 60-degree angle (Carty, 2009, p. 283). Listening from the ideal vantage point allows a realistic spatial scene as it creates a perceptual virtual sound source between the speakers, i.e. phantom image, which enhances the width of the direct sound signal from the loudspeakers (Roginska & Geluso, 2018, p. 66).

It may be inferred that the naturalness of the stereophonic sound image can be enhanced considering the psychoacoustic principles of spatial hearing, creating some realism in the presentation of direction, distance, and space in the simulated plane between the loudspeakers. The only situation in which the two-channel stereo presentation of direction, distance, and space is feasible is as a scene of spatial perspective in the simulation plane between the loudspeakers. This creates a natural two-dimensional representation of three-dimensional space (Theile, 1991, p. 767).

However, the stereo format imposes certain limitations. The phantom images are not represented behind or above the listener, and shifting away from the sweet spot cause disruption in perceived spaciousness as binaural cues are compromised (Carty, 2009, p. 283) with shifting phantom image that causes localisation error (Kendall, 2010, p. 233). In addition, the listening room's size, architecture, and acoustic treatment significantly impact the perception of stereo images; non-reverberant rooms are recommended for such reproduction systems to avoid unwanted added reverberation (Roginska & Geluso, 2018, p. 64).

Surround sound is an umbrella term to describe various loudspeaker configurations emitting sounds around the listener from multiple directions, and it is aimed at creating *"convincing spatial immersion"* (Roginska & Geluso, 2018, p. 221). This term encompasses a great variety of configurations, such as four-channel format, 5.1- and 7.1-channel surround.

Surround sound configurations have a higher capacity to create immersive effect since they allow higher resolution of phantom images between the loudspeakers; however, similar to the simpler stereo systems, the precedence effect will arise when the listener is moving out of the sweet spot, that is restricted equidistantly from all the speakers (Kendall, 2010, p. 233).

2.2 Binaural Technique

Binaural technique refers to the method of capturing three-dimensional sound scenes with microphones placed so as to mimic the human ear position and the acoustics of the pinna. This allows to capture the localisation cues (ITD, IID) and additional spectral information and encode the data into the Head-Related Transfer Function to model the acoustic behaviour of a source at a particular location in space (Roginska & Geluso, 2017, p. 88). Head Related Transfer Functions (HRTFs) describe the alteration of a sound from a specific location when it arrives at the eardrum and allows artificial spatialisation through headphones (Carty, 2009, p. 288).

One of the drawbacks of the binaural technique is considered to be the costly and time-consuming process of measuring individualised HRTF, which is a crucial element in 3D simulation. However, several methods exist to obtain the HRTF mentioned by the author, namely wave equation simulations, database matching techniques, and reciprocal HRTF measurements (Roginska & Geluso, 2017, p. 54).

2.3 Virtual Surround Sound

The field of spatial sound where spatial immersion is given great attention is the creation of virtual auditory spaces for virtual reality simulations and augmented reality. Such applications require more height, depth, and cognition data to create convincing immersive sound environments (Roginska & Geluso, 2018, p. 221). Binaural sound is an effective technique, widely used in such virtual environments, allowing effective mimicking of "human localisation cues" through independent head-related signals (Rumsey, 2002, p. 652).

The binaural headphone capabilities can be extended to create virtual listening environments through acoustic or hardware augmentation and through signal processing. The method combines the signal at the eardrum (HRTF) with the characteristics of multiple sources (loudspeakers) and with the room acoustics to create virtual phantom images and achieve a realistic experience of listening to the sound in a virtual environment (Roginska & Geluso, 2018, p. 93).

Advances in digital signal processing have rekindled interest in binaural technology. Utilising digital representations of auditory responses, it is possible to mix and pan binaurally using several audio recordings. Binaural material may be handled more easily for loudspeaker reproduction, and computer sound cards and consumer TVs employ various approaches for spatial amplification of sound from only two loudspeakers. Such gadgets usually only work in a limited range of listening environments. Head tracking is used to integrate head movements in processing, which benefits VR and gaming (Rumsey & Safari, 2012, p. 14).

2.4 Wave Field Synthesis

Wave Field Synthesis (WFS) uses a discrete array of loudspeakers to recreate the original wavefront, with sounded sources placed at any location and distance of the array, which eliminates the idea of the “sweet spot” (Carty, 2009, p. 286). While binaural, stereo and surround systems described above manipulate sound objects and auditory events at the perceptual level, the approach of sound field synthesis is capable of accurate reproduction of "temporal and spatial properties of the virtual sound sources” (Roginska & Geluso, 2018, p. 311) in a sound field that is interpreted by the auditory system (Roginska & Geluso, 2018, p. 276).

Particular limitations of the WFS, however, still lie in the fact that a large number of loudspeakers require significant computational power and complicated logistics. Technical advancements in spatial audio have long surpassed the consideration of reproduction of compelling real-world sound images. At this point, it may be assumed that the reproduction systems are of fine standards. The most recent developments in sound technology research is now increasingly focused on augmented reality systems, “which use spatial auditory content to represent information beyond that encountered in the real world” (Roginska & Geluso, 2018, p. 57). Current trends in spatial audio revolve around Extended Reality applications in the fields of healthcare, aerospace, architecture, and many more (Kailas & Tiwari, 2021). Successful implementation of spatial sound on such a large scale requires comprehensive and reliable research data in acoustics, cognition, and the principle of auditory processing. The aim of the present review is to provide a comprehensive overview of the current state of research in spatial audio and its application in augmented reality systems. The review discusses the main challenges of spatial sound research including challenges related to perception, cognition, acoustics and signal processing.

CHAPTER III

SPATIAL AUDIO QUALITY ASSESSMENT

Multichannel audio for reproduction and virtual sound environment is increasingly gaining more application not only in the domain of music and entertainment but also in the field of psychoacoustic research, which is concerned with the precise simulation of three-dimensional auditory scenes for the needs of cognitive and psycho-physiological research. This brings up a concern about how complete and convincing the reproduced sound scene is and how particular spatial attributes of a reproduced sound scene affect human perception (Guastavino & Katz, 2004, p. 1105).

Starting the discussion with spatial quality, which is categorised under other subsets such as timbral quality, technical quality and other miscellaneous qualities, and is a subset of audio quality evaluation which relates to "the three-dimensional nature of the sound sources and their environments" (Berg & Rumsey, 1999)(Berg & Rumsey, 2003a, p. 3) while other scientists elaborate on the three-dimensional aspect of it, namely emphasising on the location, dimensions, and other physical parameters of a sound source and the acoustic environment in which the source is located" (Mason and Rumsey, 2000) or "the locatedness and the localisation of the sound, how enveloping it is its naturalness and depth" (Zacharov and Nuopaniemi, 1999).

Keeping in mind the discussion mentioned above, it is clear that spatial quality evaluation deals with “sensations", i.e. "perceptual constructs that relate to the sensations of directionality, size, depth and width of reproduced sources, groups of sources and acoustical environments" (Berg & Rumsey, 2003a, p. 3) while the sound quality research examines mechanisms in listeners’ preferences and the effect of complex sound stimuli on assessment of auditory perception (Choisel & Wickelmaier, 2006, pp. 1–2). The perception of auditory quality of multichannel sound is influenced by a large number of factors. Comparing specific characteristics to overall preference and physical sound field measurements facilitates comprehension of sound quality evaluation criteria (Choisel & Wickelmaier, 2006, p. 1).

Typically, sound quality is studied in the form of mean opinion score (MOS) that consolidates all contributing perceptual aspects into a single rating. The test includes both preferential and descriptive characteristics and is used to evaluate timbral and spatial fidelity of the audio quality in multichannel reproductive systems (Rumsey, Zieliński, Kassier, & Bech, 2005, p. 968) (Rumsey, 2002, p. 651). MOS is commonly used for audio quality testing of loudspeakers, predominantly with timbral and distortion issues; however, the results of the method are considered to be limited regarding subjective evaluation of multiple spatial attributes due to its uni-scalar output (Zacharov & Koivuniemi, 2001, p. 1). Spatial attribute evaluation, however, has been in research interest in the field of room and concert hall acoustics and is often referred to as objective evaluation of sound quality. To understand more clearly, it is pertinent to distinguish between objective and subjective evaluation of sound (Griesinger, 1997), (Beranek, 1992) (Morimoto & Iida, 1995). Berg and Rumsey (2003) clearly distinguish between objective and subjective evaluation of sound quality. From objective evaluation of sound quality, they understand evaluation methods that employ "measures of physical quantities of the sound, such as sound pressure level, total harmonic distortion, reverberation time" (Berg & Rumsey, 2003a, p. 2). Subjective evaluation, however, involves perceptual judgements of the sound stimuli by the listener (Berg & Rumsey, 2003a, p. 2).

In the existing literature, sound quality is distinguished from sound character. Letowski describes the former as a “degree of satisfaction with the sound” and is mainly used by sound engineers, while the latter is a way to "differentiate between one sound from another" and lies rather in the domain of timbre and tone colour, used by musicians (Letowski, 1998, p. 1). Other researchers, Nunally and Bernstein, similarly call them *sentiments and judgments* (Nunnally & Bernstein, 1994). According to Letowski, sound quality is “that assessment of auditory image in terms of which the listener can express satisfaction or dissatisfaction with that image. Sound quality can be judged by comparing images produced by several external stimuli or referencing a perceived image to the concept residing in the listener's memory" (Letowski, 1989, p. 6). His simple yet exquisite binary model of audio quality assessment is comprised of two categories; timbral quality and spatial quality (Letowski, 1989, p. 6).

The increasing demand for spatial audio quality evaluation methods led researchers to isolate the aspects contributing to the perception of space in the course of subject tests (Berg & Rumsey, 1999; Guastavino & Katz, 2004; Mason, 2002; Rumsey, 1998; Zacharov & Koivuniemi, 2001). Further, we will review proposed sets of subjective spatial attributes within the schema-based paradigm of Rumsey (Rumsey, 2002).

3.1 Scene-Based Paradigm of Spatial Quality Evaluation

Rumsey (Rumsey, 2002) proposes a method to evaluate complex auditory scenes through a paradigm of scenes, which presupposes that the spatial attributes of the reproduced sounds are to be grouped according to their function. He mentions that the paradigm does not include the dimension of height; however other researchers found it reasonable to add it in order to extend the paradigm to three-dimensional. The addition of height appears to be reasonable with regard to the three-dimensional concept; however, this parameter has not yet been thoroughly investigated in experiments with subjects.

3.1.1 Attributes

Before moving on any further, it is necessary to discuss what exactly is meant by the term "spatial attribute" in the context of the evaluation of the sound quality. Attributes are defined as “the descriptors used for symbolising sensations or perceptual constructs” (Rumsey, 1998) and “concerns some particular feature of objects” (Nunnally & Bernstein, 1994). Thus, spatial attributes include a set of "interpretational constructs" to describe multidimensional relationships, similarities and differences between sound stimuli (Berg & Rumsey, 1999, p. 51).

Rumsey points out the necessity of distinguishing *attributes of space* from *spatial* *attributes*. The attributes of space are those attributes that the field of concert hall acoustics is mainly concerned about. Beranek outlines seven acoustical features that must be taken into consideration for concert hall design that concerns, among other, reverberation time, delay, and warmth (Beranek, 1992). Even though some of the attributes researched could be related to the perceptual spatial attributes, the main focus of the investigation is mainly on the natural acoustics, i.e. the *environment*. While, spatial attributes are concerned with "those perceptual constructs that relate to directionality, size (height), depth, and width of reproduced sources, groups of sources, and acoustical environments" (Rumsey, 2002, p. 655). Reverting back to attributes of Rumsey. The initial set-up of attributes of Rumsey includes dimensional attributes such as *distance*, *width* and *depth*. The second group are the immersive attributes such as presence and envelopment. The initial set was meant for the evaluation of fixed sound scenes.

In their following research, Berg and Rumsey introduced another attribute, calling it attribute of motion (Berg & Rumsey, 2003a, p. 2).

Recently, Kendall prosed the extended set of general attributes (Kendall, 2010). Which are as follows:

Dimensional attributes:

Direction

Distance

Extent: depth, width and height

Immersive attributes:

Presence

Envelopment

Before proceeding to the scene-based paradigm, the general notions of the aforementioned attributes need to be discussed and elaborated.

Direction

Direction is also referred to as *localisation* (Berg & Rumsey, 2003b), *sense of direction* (Zacharov & Koivuniemi, 2001), or *definition of sound images* (F. Toole, 1997). The definition of direction, given in the final evaluation experiment, is "how easy it is to perceive a distinct location of the source – how easy it is to pinpoint the direction of the sound source. Its opposite is when the source's position is hard to determine – a blurred position" *(Berg & Rumsey, 2003a, p. 12).*

Distance/Depth

Koivuniemi proposed the definitions for distance and depth that were developed for verbal protocol analysis of spatial attribute elicitation (Zacharov & Koivuniemi, 2001, p. 7):

> "Distance to events: This attribute simply describes the actual distance where the sound events appear to originate. A positive value implies that the sound sources are sensed to be far from the listening point."
>
> "Sense of depth: Describes how strongly the sensation of distance is perceived, or how ambiguous the sensation of distance is. Once again this

> assesses whether several sound events can be discriminated in terms of distance. A negative value could mean that distances for all events are ambiguous except those originating from the transducer's position."

The definition of distance and depth in the scene-based paradigm is reviewed independently with respect to the levels of organisation. Distance refers to the perceived range between the listener and the sound source, while depth is attributed to the sense of perspective in the scene (Berg & Rumsey, 2003b, p. 660). Details will be provided below.

Width

Width might appear to be an ambiguous term to be applied generally; therefore, it will be reviewed later with application to the grouping levels of the scene-based paradigm. However, Koivuniemi and Zacharov proposed a somewhat cover-all definition for the needs of the experiment. Their equivalent to *width* of Rumsey is *broadness* which describes "how wide an area the perceived sound event seems to have". A strong sense of broadness would occur when the listener is *enveloped* by sound coming from all directions (Zacharov & Koivuniemi, 2001, p. 7).

Presence

Presence, or *sense of space* (Zacharov & Koivuniemi, 2001), refers to the experience of "being in the same acoustical environment as the sound source" (Berg & Rumsey, 2003b, p. 12).

Envelopment

Envelopment is interrelated with the notion of width. In the field of live and reproduced sound, it might as well be termed as "spaciousness", "spatial impression", or "ambience". However, it appears to be a broader term. In concert hall acoustics, envelopment is comprised of the *apparent source width* and *listener envelopment* (Berg & Rumsey, 2003a, p. 4). Therefore, "source envelopment can be understood as the extent to which the sound source envelops/surrounds/exists around the listener. The feeling of being surrounded by the sound source. If several sound sources occur in the sound excerpt, sound envelopment assesses the sound source perceived to be the most enveloping" (Berg & Rumsey, 2003a, p. 12).

3.2 Individual Source, Ensemble, Environment

Rumsey proposes that attributes in a spatial auditory scene can be grouped into micro and macro entities according to their function. A typical scene consists of *individual sources* located within an acoustic environment. Individual sources can have attributes of width, distance, direction, and possibly depth. The sources may be cognitively grouped into one or more macro entities, so-called *ensembles*. The ensembles are attributed to lateral location, width, depth, and distance. There may exist a number of ensembles, as well as they can be included in bigger ensemble agglomerations. The *acoustic environment* that contains the ensembles is the next perceptual plane and can have all the aforementioned attributes with additional attributes of envelopment and presence. The resulting set of attributes is as follows:

Dimensional attributes:

Direction — source, ensemble, environment

Distance — source, width, environment

Depth/width/height — source, width, environment

Immersive attributes:

Envelopment — ensemble, environment

Presence — source, ensemble, environment

The main attributes of micro and macro groupings will be reviewed in the following sub-section.

3.2.1 Attributes of Individual Sources

The attributes of individual sources cannot be understood without clearly stating the sub-attributes, which are discussed below.

Distance

As mentioned before, sound source distance is the perceived range between the listener and the reproduced source. Distance may as well refer to the source location; however, the accuracy of perceptual positioning can be obstructed within a reverberant environment (Rumsey, 2002, p. 657). In such an enclosed environment, the distance cues to the sound source are provided by the ratio between direct and indirect sound. The perception of distance tends to be relatively inaccurate with unfamiliar sources and with the sources positioned either too close or too far away from the listener (as cited in H. A. Lynch, 2014, p. 63).

Width

Individual source width (ISW) refers to the perceptual extent or angle occupied by a single sound source. Concert hall acoustic describes the phenomenon of apparent source width (ASW) in association with a sound source that is comprised of the level, direction, frequency content, and character of early reflections (Rumsey, 2002, p. 658). In evaluation experiments, the subjects are asked to disregard the width of the environmental reverberation and to focus solely on the extent of the source (Berg & Rumsey, 2003a, p. 12).

Depth

The perception of depth in individual sources has not found substantial approval in subjective experiments; however, some researchers (as cited in Rumsey, 2002, p. 660) found that some subjects were able to distinguish between flat and curved sources.

3.2.2 Attributes of Ensembles

The attributes of ensembles comprise sub-attributes which are discussed below.

Width

Ensemble width is "the perceived width/broadness of the ensemble, from its left flank to its right flank. The angle occupied by the ensemble. The meaning of *the ensemble* is all of the individual sound sources considered together" (Berg & Rumsey, 2003, p. 12). Unlike the ISW, ensemble width estimation cannot rely on the ratio between direct and indirect sound and is rather perceived cognitively. (Rumsey, 2002, p. 658) Toole (F. E. Toole, 2008, p. 99) refers to it as the *width of the sound stage*.

Depth

Ensemble depth was proved to be more apparent for estimation, relating to the perception of front-back dimensions (Rumsey, 2002, p. 660). In reproduced sound, the perception of depth is provided by the amount and quality of indirect sound rather than the direct one or the relative loudness of the ensemble (Theile, 2001, p. 21).

3.3 Attributes of the Environment

The attributes of environment require clarification on sub-attributes, which are discussed below:

Width

The attribute of the environment width is closely related to the notion of *spaciousness* and relies on cognitive separation of individual sources from the background spatial impression. It is also related to the sense of presence discussed above, i.e. the impression of difference between a narrow and wide space (Rumsey, 2002, p. 659).

Depth

The elicitation experiments are said not to have found reliable evidence for a separate construct of environment depth. Overall, the perceptual construct of width seemed to be dominant for reasons yet to be explained (Rumsey, 2002, p. 660).

3.4 Envelopment, Spaciousness, and Spatial Impression

Unlike the linear attributes of dimension, envelopment, spaciousness, and spatial impression represent rather semiabstract and multidimensional impressions and are harder to elicit in uniform experiments. Moreover, the terms are subject to multiple interpretations across the literature. Generally, the terms are used to describe the sensation of immersion within a sound field (Rumsey, 2002, p. 661).

In concert hall acoustics, the term envelopment was attributed to the “feeling of surroundedness or engulfment in a well-designed concert hall”, which was achieved by making the reflected sound arrive from the sides after the direct sound (as cited in Roginska & Geluso, 2018, p. 384), while Toole (F. E. Toole, 2008, p. 34) defines spaciousness or spatial impression as a sensation of listening to sound in a large space and devises two components to it:

- Apparent source width (ASW) — an auditory illusion of the sound source being wider than the visible source, and

- Listener envelopment (LEV) — a sense of being surrounded by diffused sound without any localisable sound sources (F. E. Toole, 2008, p. 34).

Letowski, on the other hand, defines spaciousness as the sense of being surrounded by a space, which is dependent on the amount, character, and location of sound sources around the listener. He agrees that the percept is described in relative terms because it does not yet have specific descriptive dimensions (Letowski & Letowski, 2012, p. 4).

Rumsey (Rumsey, 1998, p. 129) draws an important connection between Bregman's idea of auditory streaming and the notion of spatial impression of Griesinger. According to Bregman, an auditory scene can be segregated into streams using, among others, location cues (via interaural time difference, ITD, and interaural intensity difference, IID) (Harding, Cooke, & König, 2007, p. 159). Griesinger, in turn, elaborates on the effect of spatial impression on stream segregation. He separates spatial impressions into three categories: Continuous Spatial Impression (CSI), Early Spatial Impression (ESI), and Background Spatial Impression (BSI). He argues that the sense of full spaciousness or envelopment (CSI) can be experienced when the correlation between the direct continuous sound and the reflected energy is so strong that the interaural fluctuation cannot separate it into separate streams. On the other hand, the sound can be segregated if the reflected energy arrives within 50ms of the direct sound. It gives rise to an impression of a small room (ESI). Reflected signal arriving more than 50ms after the direct sound results in BSI, i.e., the impression of a large space with diffused energy and provides a good impression of envelopment (as cited in Rumsey, 1998, p. 129).

The discussion about perceptual attributes of space in reproduced audio is far from reaching somewhat congruent conclusions. The disparity lies in a wide variety of applied methods of subjective elicitation of spatial attributes as well as semantic descriptors for such experiments. The authors also point out the need to extend the methodological instruments to be applicable to dynamic scenes. However, it can be agreed that there exist common perceptual patterns that can be expressed in attributive descriptors and elicited. The scene-based paradigm of Rumsey is a significant contribution to the existing methods of obtaining and structuring of subjective attributes. The main contribution of this paradigm is a theoretical framework for the description and analysis of space-related scenes. It provides an essential tool for the study of visually perceived environments. The method can be used to structure scene descriptions, which are then used as the basis for further research on spatial cognition.

CONCLUSION

The auditory system is thought to apply a hierarchy of computations that transform low-level features such as amplitude and frequency into perceptually significant information. The hierarchy consists of detection, grouping, and segregation at the lowest level; feature analysis in the middle layer; and finally, recognition of whole objects at the top. The hierarchy of auditory processing is an ordered sequence of stages that transform the incoming signal into a meaningful percept. This hierarchy has been reviewed to include feature extraction, pattern recognition, and object identification at different levels in the auditory system. The latest developments in sound technology are now focused on the creation of augmented reality systems, which use spatial auditory content to represent information beyond that found in everyday life.

The current paper provided a review of the current literature on auditory quality assessment, with a focus on the hierarchical organisation of this system. The review also discussed the current developments in spatial sound technologies and their idiosyncrasies.

References

Andringa, T. (2011). *Audition: From sound to sounds*. https://doi.org/10.4018/978-1-61520-919-4.CH004

Baalman, Marije A.J. (2010). Spatial Composition Techniques and Sound Spatialisation Technologies. *Organised Sound*, *15*(03), 209–218. https://doi.org/10.1017/S1355771810000245

Baalman, Marije Alberdina Johanna. (2008). *On Wave Field Synthesis and electro-acoustic music, with a particular focus on the reproduction of arbitrarily shaped sound sources*. https://doi.org/10.14279/DEPOSITONCE-1883

Barreiro, D. L. (2010). Considerations on the Handling of Space in Multichannel Electroacoustic Works*. *Organised Sound*, *15*, 290–296. https://doi.org/10.1017/S1355771810000294

Barrett, N. (2002). Spatio-musical composition strategies. *Organised Sound*, *7*, 313–323. https://doi.org/10.1017/S1355771802003114

Bates, E. (2009). *The composition and performance of spatial music*.

Benoudjit, M. A. (2012). *Human perception and space classification: The Perceptive Network*. Retrieved from https://semanticscholar.org/paper/a756c4579247047ffc9f14890497ee9ce9409c55

Beranek, L. L. (1992). Concert hall acoustics—1992. *The Journal of the Acoustical Society of America*, *92*(1), 1–39. https://doi.org/10.1121/1.404283

Berg, J., & Rumsey, F. (1999). *Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods*. Retrieved from https://semanticscholar.org/paper/9d635f51f74d9ce57d12d88098d28ef6ac3a524a

Berg, J., & Rumsey, F. (2000a). In Search of the Spatial Dimensions of Reproduced Sound: Verbal Protocol Analysis and Cluster Analysis of Scaled Verbal Descriptors. *Journal of The Audio Engineering Society*. Retrieved from https://semanticscholar.org/paper/bbe7486d290d510150812a9d471a0451ecd749ab

Berg, J., & Rumsey, F. (2000b). In Search of the Spatial Dimensions of Reproduced Sound: Verbal Protocol Analysis and Cluster Analysis of Scaled Verbal Descriptors.

Berg, J., & Rumsey, F. (2003). *Systematic Evaluation of Perceived Spatial Quality*. Retrieved from https://semanticscholar.org/paper/2f3258bf20f463637b5b35d72cb8a0d82d68b839

Boren, B., & Musick, M. (2018). *Spatial Organization in Musical Form*. Retrieved from https://semanticscholar.org/paper/f89e45d7ca6bed6b295f474a77bf2644ae6d9ca0

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, Mass.: MIT Press. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1658

Choisel, S., & Wickelmaier, F. (2006a). Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound. *Journal of The Audio Engineering Society*, *54*, 815–826.

Choisel, S., & Wickelmaier, F. (2006b). *Relating auditory attributes of multichannel sound to preference and to physical parameters*. Retrieved from https://semanticscholar.org/paper/35d8dc4eceef62f76d08dd52b3806dde53eb08ba

Choisel, S., & Wickelmaier, F. (2007). Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. *The Journal of the Acoustical Society of America*, *121 1*, 388–400. https://doi.org/10.1121/1.2385043

Coates, P., Derix, C., Lau, T., Parvin, T., & Puusepp, R. (2005). *Topological Approximations for Spatial Representations*. Retrieved from https://semanticscholar.org/paper/0469584fa69bd266759f3056fc757e3f15d0e773

Cohen, M., & Martens, W. L. (2020). *Spatial Soundscape Superposition and Multimodal Interaction*. https://doi.org/10.1007/978-3-030-00386-9_13

Cook, P. R. (1999). *Music, cognition, and computerized sound: An introduction to psychoacoustics*. Cambridge, Mass.: MIT Press. Retrieved from http://site.ebrary.com/id/10229599

Derix, C. (2006). *Approximating Phenomenological Space*. https://doi.org/10.1007/11888598_15

Francombe, J., Brookes, T., & Mason, R. (2015). *PERCEPTUAL EVALUATION OF SPATIAL AUDIO : WHERE NEXT ?* Retrieved from https://semanticscholar.org/paper/652855436ecaa361b9e07195b69072e3a702760b

Gelfand, S. A. (2018). *Hearing: An introduction to psychological and physiological acoustics* (Sixth edition). Boca Raton: CRC Press.

Gilkey, R. H., & Anderson, T. R. (2014). *Binaural and spatial hearing in real and virtual environments*. Retrieved from http://site.ebrary.com/id/10843472

Grigoropoulos, G. (2016). *Identity spaces music space as a medium for sound*. Retrieved from https://semanticscholar.org/paper/6b424a4a3448ea8a4c69a30e751cccddd0582afb

Guastavino, C., & Katz, B. F. G. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *The Journal of the Acoustical Society of America*, *116 2*, 1105–1115. https://doi.org/10.1121/1.1763973

Hagan, K. L. (2008). *Textural composition and its space*. Retrieved from https://semanticscholar.org/paper/38c352e323e43372dde49a411da8f069957abd0f

Hagan, K. L. (2017). Textural Composition: Aesthetics, Techniques, and Spatialization for High-Density Loudspeaker Arrays. *Computer Music Journal*, *41*, 34–45. https://doi.org/10.1162/COMJ_a_00395

Haggard, M. P. (1982). Hearing: An Introduction to Psychological and Physiological Acoustics. *Journal of Neurology, Neurosurgery & Psychiatry*, *45*, 1175–1175. https://doi.org/10.1136/jnnp.45.12.1175-b

Harding, S., Cooke, M., & König, P. (2007). *Auditory Gist Perception: An Alternative to Attentional Selection of Auditory Streams?* https://doi.org/10.1007/978-3-540-77343-6_26

Harley, M. A. (2016). *Space and Spatialization in Contemporary Music*.

Holbrook, U. A. S. (2019). Sound Objects and Spatial Morphologies. *Organised*

Sound, *24*(1), 20–29. https://doi.org/10.1017/S1355771819000037

Jean-Marc. (2007). *EFFICIENT DESCRIPTION AND RENDERING OF COMPLEX INTERACTIVE ACOUSTIC SCENES*. Retrieved from https://semanticscholar.org/paper/037c22ae8b8af8024c0b17aa9aeb213755b84ab8

Johnson, E., & Adamo, N. (2010). A Study of the Effects of Immersion on Short-term Spatial Memory. *World Academy of Science, Engineering and Technology*, *71*, 582–587.

Kendall, G. (2008). *WHAT IS AN EVENT? The EVENT Schema, Circumstances, Metaphor And Gist*.

Kendall, G. S., & Cabrera, A. (2011). *Why things don't work: What you need to know about spatial audio*. Retrieved from https://semanticscholar.org/paper/9a01139b293bb52215aa16a27d1b50e5b72c2678

Kendall, G. s. (2010). Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music. *Organised Sound*, *15*, 228–238. https://doi.org/10.1017/S1355771810000336

Khosravi, P. (2012). *Spectral spatiality in the acousmatic listening context*. City University London. Retrieved from https://openaccess.city.ac.uk/id/eprint/2717/

Kronland-Martinet, R., Ystad, S., & Jensen, K. (2008). *The Artistic Play of Spatial Organization: Spatial Attributes , Scene Analysis and Auditory Spatial Schemata*. Retrieved from https://semanticscholar.org/paper/0667c9db57f2462980f8112032b5299623fe6649

Langley, P., Derix, C., & Coates, P. (2007). *Meta-Cognitive Mappings: Growing Neural Networks for Generative Urbanism*. Retrieved from https://semanticscholar.org/paper/7ff619e3ad17636af81d4a1b3b81503d5a14a582

Letowski, T. (1989). Sound Quality Assessment: Concepts and Criteria. *Journal of The Audio Engineering Society*. Retrieved from https://semanticscholar.org/paper/7d8287af5caec10fae3da5e40a122e77d116f03d

Letowski, T. (1998). Perceived sound quality: Back to the future. *Proceedings of the NVH Symposium*, 71–76.

Lynch, H. A. (2014). *Space in multi-channel electroacoustic music: Developing sound spatialisation techniques for composing multi-channel electroacoustic music with emphasis on spatial attribute perception*. Retrieved from https://semanticscholar.org/paper/3b58d262b62141c34b733c1d4cc667d1cef77e9f

Lynch, H., & Sazdov, R. (2011). *An Investigation Into Compositional Techniques Utilized For The Three- Dimensional Spatialization Of Electroacoustic Music*. Retrieved from https://semanticscholar.org/paper/30bdd8f3c5b9a1ff259134de66fd50807cccaf14

Lynch, H., & Sazdov, R. (2013). *An Investigation into the perception of Spatial Techniques used in Multi-channel Electroacoustic Music*. Retrieved from https://semanticscholar.org/paper/fd7390095222d7381635847edcc3c79c69834b61

Lynch, H., & Sazdov, R. (2017). A Perceptual Investigation into Spatialization

Techniques Used in Multichannel Electroacoustic Music for Envelopment and Engulfment. *Computer Music Journal*, *41*, 13–33. https://doi.org/10.1162/COMJ_a_00401

Macedo, F. (2011). *PHENOMENOLOGY, SPATIAL MUSIC AND THE COMPOSER: PRELUDE TO A PHENOMENOLOGY OF SPACE IN ELECTROACOUSTIC MUSIC*.

Macedo, F. (2015). Investigating Sound in Space: Five meanings of space in music and sound art. *Organised Sound*, *20*(2), 241–248. https://doi.org/10.1017/S1355771815000126

Martens, W. L., & Cohen, M. (2020). Spatial soundscape superposition, Part I: Subject motion and scene sensibility. *Acoustical Science and Technology*, *41*, 288–296. https://doi.org/10.1250/ast.41.288

Mason, R. (2002). *Elicitation and measurement of auditory spatial attributes in reproduced sound*. Retrieved from https://semanticscholar.org/paper/acda853a5535fa67b5c1ccde5262f8af968f2db6

Mason, R., Ford, N., Rumsey, F., & Bruyn, B. D. (2001). Verbal and non-verbal elicitation techniques in the subjective assessment of spatial sound reproduction. *Journal of The Audio Engineering Society*, *49*, 366–384.

Memory, Spatial. (n.d.). In *Encyclopedia of the Neurological Sciences*.

Negrão, M. C. (2014). *ImmLib—A new library for immersive spatial composition*. Retrieved from https://semanticscholar.org/paper/b4c3210e4acc60d20ee1a172ef42b956577e5399

Normandeau, R. (2009). Timbre Spatialisation: The medium is the space. *Organised*

Sound, *14*, 277–285. https://doi.org/10.1017/S1355771809990094

Nudds, M., & O'Callaghan, C. (Eds.). (2009). *Sounds and perception: New philosophical essays*. Oxford ; New York: Oxford University Press.

Nystrom. (2016). *Low-level topology of spatial texture*. Retrieved from https://semanticscholar.org/paper/23959e2753ae6448cb929d06a9f0717d100c7835

Nyström, E. (2013). *Topology of spatial texture in the acousmatic medium*. Retrieved from https://semanticscholar.org/paper/b060984439a8f21e916d24fa9fb0ab429f132dbb

Nyström, E. (2018). *Topographic Synthesis: Parameter Distribution in Spatial Texture*. Retrieved from https://semanticscholar.org/paper/13066d0369802c69b48ad646f3bcaf3b8cccfefe

Nyström, E., & Kingdom, U. (2010). *Elemental Chemistry Textons and a Vision of Space in Acousmatic Music*. Retrieved from https://semanticscholar.org/paper/82ff334c941efa3b7f58ebb0c111f4a081cdd16b

Paine, G., Sazdov, R., & Stevens, K. (2007). *Perceptual Investigation into Envelopement, Spatial Clarity, and Engulfment in Reproduced Multi-Channel Audio*. Retrieved from https://semanticscholar.org/paper/c904cf313a0f4a27ebbcfcc6ef358edc97559b45

Peters, N., Marentakis, G., & McAdams, S. (2011). Current Technologies and Compositional Practices for Spatialization: A Qualitative and Quantitative

Analysis. *Computer Music Journal*, *35*, 10–27. https://doi.org/10.1162/COMJ_a_00037

Pouru, L., & Oliva, D. (2019). *THE PARAMETERS OF REALISTIC SPATIAL AUDIO*. Retrieved from https://semanticscholar.org/paper/6fe2a59e57a7c1296786104ad3c40aaaa00a312c

Psarra, S. (2003). *Top-down and bottom-up characterisations of shape and space*. Retrieved from https://semanticscholar.org/paper/ccf5f5a4c4b7bcad72cb0151763dd364d83b0055

Psarra, S., & Grajewski, T. (2001). *Describing Shape and Shape Complexity Using Local Properties*. Retrieved from https://semanticscholar.org/paper/d5254d0a90f9563326fe916a8459e110441404db

Ratti, F. S., & Bravo, C. F. (2017). Space–Emotion in Acousmatic Music. *Organised Sound*, *22*, 394–405. https://doi.org/10.1017/S1355771817000449

Ricard, J. (2004). *Towards computational morphological description of sound*.

Roginska, A., & Geluso, P. (Eds.). (2018). *Immersive sound: The art and science of binaural and multi-channel audio*. New York ; London: Routledge, Taylor & Francis Group.

Romigh, G. D., & Simpson, B. D. (2014). Do you hear where I hear?: Isolating the individualized sound localization cues. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00370

Ruff, R. M., & Perret, E. (1976). Auditory spatial pattern perception aided by visual choices. *Psychological Research*, *38*, 369–377.

https://doi.org/10.1007/BF00309042

Ruff, R. M., & Perret, E. (1982). Spatial Mapping of Two-Dimensional Sound Patterns Presented Sequentially. *Perceptual and Motor Skills*, *55*, 155–163. https://doi.org/10.2466/pms.1982.55.1.155

Rumsey, F. (1998). *Subjective Assessment of the Spatial Attributes of Reproduced Sound*. Retrieved from https://semanticscholar.org/paper/c40c01988169200a95937af3fff62ef9cf48de47

Rumsey, F. (2002). Spatial quality evaluation for reproduced sound: Terminology, meaning and a scene-based paradigm. *Advances in Engineering Software*. Retrieved from https://semanticscholar.org/paper/eae86d8e60d3c9440c4a702dfe8194541f81bacd

Rumsey, F. (2006). *Spatial audio and sensory evaluation techniques—Context, history and aims*. Retrieved from https://semanticscholar.org/paper/2e553b1daceecc6e94db1e612e76af24e66ce4ea

Rumsey, F., & Berg, J. (2001). *Verification and correlation of attributes used for describing the spatial quality of reproduced sound*. Retrieved from https://semanticscholar.org/paper/f762c014aca13e1a68f38b3f787ed5c49cc2b173

Rumsey, F., & Safari, an O. M. C. (2012). *Spatial Audio*.

Rumsey, F., Zieliński, S., Kassier, R., & Bech, S. (2005). On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America*, *118 2*, 968–976.

https://doi.org/10.1121/1.1945368

Seaback, R. (2020). Anacoustic Modes of Sound Construction and the Semiotics of Virtuality. *Organised Sound*, *25*, 4–14. https://doi.org/10.1017/S1355771819000414

Seddon, A. (2013). *Recurrence in acousmatic music*.

Shinn-Cunningham, B. G., & Wang, D. (2008). Influences of auditory object formation on phonemic restoration. *The Journal of the Acoustical Society of America*, *123 1*, 295–301. https://doi.org/10.1121/1.2804701

Suzuki, Y., Brungart, D., Iwaya, Y., Iida, K., Cabrera, D., & Kato, H. (2011). *Principles and Applications of Spatial Hearing*. WORLD SCIENTIFIC. https://doi.org/10.1142/7674

Theile, G. (1991). On the Naturalness of Two-Channel Stereo Sound. *Journal of The Audio Engineering Society*, *39*, 761–767.

Trowitzsch, I. (2020). *Robust sound event detection in binaural computational auditory scene analysis*. https://doi.org/10.14279/DEPOSITONCE-9857

Young, J. (2004). Sound morphology and the articulation of structure in electroacoustic music. *Organised Sound*, *9*(1), 7–14. https://doi.org/10.1017/S1355771804000032

Zacharov, N., & Koivuniemi, K. (2001a). *Audio descriptive analysis & mapping of spatial sound displays*. Retrieved from https://semanticscholar.org/paper/f2dd48bdde608cf97663f497901c67bbf231b8a4

Zacharov, N., & Koivuniemi, K. (2001b). Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis and Listener Training. *Journal of The Audio Engineering Society*. Retrieved from https://semanticscholar.org/paper/a71579b636b3ea962c8dbb108a2b94d5bdd bf278