Fakultät Design, Medien und Information Department Medientechnik

An Installation using Eurorack Modular Synthesis in Four-Channel Format

Hadi Abilmona

Hamburg University of Applied Sciences

Fakultät Design, Medien und Information

Masterthesis

Prof. Thomas Görne

Stefan Troschka

June 15, 2022

Table of Contents

Abbreviations

The following abbreviations were used in this paper:

Abbreviation	Description
1v/O	One Volt per Octave
AD/AR	Attack, Decay/Attack, Release
ADSR	Attack, Decay, Sustain, Release
AM	Amplitude Modulation
CV	Control Voltage
FM	Linear Frequency Modulation
FM EXP	Exponential Frequency Modulation
LFO	Low Frequency Oscillator
PW/PWM	Pulse Width/Pulse Width Modulation
TZFM	Through Zero Frequency Modulation
VCA	Voltage Controlled Amplifier
VCF	Voltage Controlled Filter
VCO	Voltage Controlled Oscillator
XFM	Cross-Modulation

Abstract

In this thesis, the use of space as a controlled parameter in a four-channel modular synthesizer installation is considered. Human perception, psycho-acoustics, and the hearing apparatus are of particular concern to spatial hearing and perception, therefore basic theories of perception and brief history are considered. The use of space as an artistic parameter generates certain effects. An approach to sound and spatialization is introduced and discussed for modular synthesizers. What are the effects of space on sound and of sound on space, and how do they influence each other? What parameters can be controlled and how can this knowledge be applied in order to create purposeful applications? Using artistic research methodology, possible answers and outcomes for these questions have been hypothesized, and examined, with the reflection on experimentation and its generated material.

Introduction

Preface

Art is a widely spread, albeit not easily defined notion. It is a form of expression, a depiction and subsequent interpretation of things, and is subordinate to them because it is reliant on them. As formulated by (Adajian, Thomas, & Edward N. Zalta 2018), "grasped perceptually, artworks present only an appearance of an appearance of the forms". But what are forms? Form is a work of symbolism that communicates in some form. Form is the realization and perception of things (Valéry 1987: 61). These definitions are nonetheless not exclusive to art, because there are other forms of expression and representations that lack formal art. As part of his larger theory of aesthetics, Immanuel Kant argues that art, although a means without an end, stimulates cognition and communicates socially. Georg Wilhelm Friedrich Hegel defined art as a solely perceived articulation of metaphysical truth, which is the universe's recognition of concept and rationality. He adds that the cultural aspect of art and its artistic statement can be compared with philosophy and religion, whereby art differs from the latter two in that it operates on the level of human senses. He also states that art is a prevalent form of cultural expression in most cultures (Adajian, Thomas, & Edward N. Zalta, 2018). Current definitions of art vary but distinguish two main aspects, cultural and transcultural artistic phenomena. A more recent trial of defining art, is that it encompasses the perception and the embodiment of the perceived. This embodiment, however, is almost always differential, i.e. depending on the perceiver, as objects are perceived uniquely by every individual, thus making art unique to the perceiver. Here, individual experiences as well as cognition play a major role in shaping this plastic perception (Badura et al., 2015). Art is to see with the eyes of the artist (Plessner, 1982, pp. 169–170), and it is conception after reflection upon the perceived. Art can hardly be separated from its technical or material properties. For example, music needs a material medium with certain characteristics in order to be manifested, as much as colors have wavelengths, as well as a metaphysical meaning. It is an externalization after the internalization of an external object or phenomenon (Blesser & Salter, 2009, p. 131). It is also a continuum, although its technical and metaphysical parts are distinctive and sequential. Stockhausen reports that:

> new means change methods, which change experience, which in turn change man, and particularly in the field of music, because the sound waves penetrate very deep into the molecular and atomic layers of ourselves. Whenever we hear sounds, we are changed, we are no longer the same. This is the case when we hear sounds organized by other human beings. (Stockhausen, 1972a)

Stockhausen contemplated how perceptual effects of art and the development of art have imminent influence on our physiology, indicating that the art at the time has changed and it challenges and redefines our senses. He went on to explain that we become what we experience, in that we expand and change. Art is self-revelatory, it uncovers us and our existence. "There is no difference between composing music and thinking of the stars and the way the whole universe is functioning" says Stockhausen, (1981).

Dieter Mersch (Badura et al., 2015, p. 49) mentions that the aesthetic of phenomena is a work of its own. It constructs a working process which may or may not include a research process. In the case where it includes research, it may then create the opportunity of research within it. The focus here, however, is on the work of art itself as a phenomenon. It is a statement of some kind, which manifests itself in a form or thing that communicates. It is not a means of finding out whether something is useful or not, but is rather an articulation of something that is self-enclosed and stands alone independently.

It is also important to distinguish between the form of art and the material it uses, as well as the form and content it carries, and between the original and the copy. It is established that a phenomenon with artistic dimension or aesthetic is a form of something perceived, reformed and then communicated again in a representational form, which can be perceived again as an original. A copy is something that resembles or symbolizes the original or a representation of it. Installation art is not replicable, since it is limited to a specific place, time, and content. It has a representational aspect which is conditioned by itself. Important to the aspect of originality is that installation art is also not collectable. It does not come in quantities or multiples. This increases the value of the experience of it, and makes an installation a formal representation of the thought in physical material, a form of the processes of the author's intentions of meaning, as well as a state achieved through imagination. In the mentioned duality of form and content as well as form and material, it should be noted that content prevails over form and material, as does an original prevail over the copy. Through the twentieth century and onwards, art became more abundant and variable in its forms, specifically so that a part of the artistic work and its perception is lent to the receiver. In this way the experiencers of an installation can move around and pick a spot or a corner that suits them and their aesthetic tendency. This carries the name of Reception theory, to cite Umberto Eco in Künstlerische Forschung (Badura et al., 2015). Similarly, Mersch (Badura et al., 2015, p. 50) says that text is not completely formulated without the reader. Helmut Plessner also points out that one must be foreign or unfamiliar to an object they want to see, otherwise they are not successful in seeing it. Acousmatic sound also suits such a claim, for only unfamiliar sounds can become known (not to mention that in sound, the opposite applies, where a sound loses its meaning or reference by being listened to several times). One must also be at a distance from something to be able to see it, as the person being the object themselves, renders it impossible for them to see the object, much like one can not look at themselves from afar (Badura et al., 2015, p. 62).

Sound installations are forms of art that originated in the 1960s. They are experienced and more so inhabited, much like creatures inhabit their surroundings. They are the opposite of a gallery, where art represents a window of the environment to be experienced, with clear cut boundaries that the perceiver can not cross. In contrast, installations allow a location to be a part of the art, and not just a space to display it, contributing to the whole that becomes greater than the sum of its parts (gestalt theory). They also transcend social traditions and allow the physical movement of the perceiver. Additionally, they can take place in a public space. This does not confine the art to a specific subcategory of people, but rather opens it up to the whole of the society. Sound installations can have rather unorthodox setups of loudspeakers, which can contribute to the uniqueness and originality of the piece. While in a studio or at a concert, great attention is paid to how the loudspeakers are positioned, installation setups allow for the adaptation of the environment to the work of art, and vice versa. The acoustic features of the space itself where the phenomenon takes place also contributes to the uniqueness of the art. As a result, this can be useful to the gestalt of things, such as setting loudspeakers as points in space or moving sounds in space (Oliveira et al., 1994). Therefore, installations are site-specific and consequently, usually hard to transport, as they are not constituted of one or more objects, but are the sum of several aspects like material, concepts, and environments. Ancient drawings in caves, like those in the Lascaux

Cave in France (to be mentioned later) are considered installations, not only because of the specific resonances found at the drawing’s locations, but also because the stone itself onto which the drawings were applied contributed to their color, as well as other factors that lead to the gestalt of the end product (Rosenthal & Rosenthal, 2003).

Installation is a term that originated in the 1960s to help differentiate between an environment and a piece of art. An environment is or houses a space where installations take place. Installations, as previously mentioned, incorporate spaces as a part of the work itself, and this applies to the installation observers, where they also become part of the work themselves (Badura et al., 2015, p. 150). It is an interaction between the object, the art, and the subject or observer. This is also true not only when an installation is interactive, but in art generally, as Wolfgang Metzger (Badura et al., 2015, p. 60) already showed in his works. He states that different individuals react differently to the same light impulses, although physiologically, the impulse was identical to all subjects. Installations can take many forms, one of which is sound. Since the second half of the twentieth century, art became no more restricted to being viewed in a painting, hung on a wall in a museum, listened to through stereo loudspeakers or at a concert hall, but rather has been opened up and the limits between the art and the viewer have vanished, bringing both into the same realm. Installations challenge the perception of the audience and prompt questioning the installation experience. They are in constant interaction with artistic research, which in turn results in shaping the installations. Changing perspective in installations happens in changing positions or locations. In sound installations, different sound projection formats yield different results. Headphones tend to project sound into our ears, which is a good way to eliminate the listening space's sonic contribution, while listening to stereophonic playback systems positions the listener in front of the sound, as two loudspeakers can only generate one horizontal plane of events contributing to less immersiveness than other (immersive) formats. Blesser & Salter (2009) state that only with a multichannel format can envelopment be achieved at its best. Furthermore, installations are mediums that can host several forms of communication. For instance, a visual installation might have television screens and screen projectors, but also physical sculptures. An audiovisual installation, on the contrary, can hold television screens and lights, as well as loudspeakers and acoustic sculptures. It can be assumed that it is universal and embeds the objects of the art in the environment to make up a unique experience. Installations are experiences that parallel everyday life experiences and are based on two pillars, space and time. As in day-to-day life, installations set a challenge for the perceiver to gain knowledge in space and over time through encountered situations. Also similar to daily life, they are perceived in a nonlinear fashion, as each moment might or might not contain information about one thing or another (Rosenthal & Rosenthal, 2003).

Sound design is a kind of art that conjoins artistic talent and technical ability. It does not exist without either and is a form of sonic art that involves both (Görne, 2017, p. 12). Electronic sound, including its production using modular synthesis, changed the course of sound history in which new, unprecedented sounds came to light, which had no reference to our evolutionary and former experiential experiences. With the approach of the twentieth century, music constituted of the twelve semitone scales of the western world started to deviate away from its traditional form and in the direction of sound design. Artists started to test, question, bend and mold the centuries old established musical styles, inspired by technological advancements and new philosophical ideologies. For instance, in the 1920s Arnold Schönberg started making music without the use of musical theory. Olivier Messiaen redefined rhythm and harmony. Pierre Schaeffer, the founder of the term "Musique Concréte" composed music by producing or recording sounds, and then manipulating them into a musical context, abandoning traditional instruments (Görne, 2017, p. 15). The latter was not only a sound design pioneer, he also integrated space into his pieces, by standardizing a four-channel setup for “Musique Concréte” and later five loudspeakers. He also made clear that the spatial attributes of his music play a definite role in its consumption, as did the movement of sounds. Especially interesting is also the disappearance of boundaries between sound design and music, where sound design could become musical and vice versa. This is a field of sound generation where the rest is left to the imagination to categorize and to the perception to interpret and organize (Görne, 2017, p. 16).

Sounds trigger emotions, ideas, concepts, and memories. As these concepts exist internally in the perceiver, they are formed by past personal experiences. Nevertheless, a large amount of the sounds are common to a large number of individuals, making referencing and more precisely conditioning (triggering the recall of the references) possible. Through already learned conditioning, it is possible to create new associations and eventually, new conditions. This base, established through the collective experience, makes it possible to deliver or communicate something without speech and body language. It was not until the mid 20th century that sound was observed in isolation, independent of the object causing it.

Emmerson (1998) claims that "we listen to any sound conditioned by our primeval past and evolution." The sound of a car passing by would be recognized almost all over the world as the movement of a physical vehicle, a case of sound being heavily associated with its object, the car (engine, wheels, brakes), in spatial and temporal motion. The movement of a car from left to right communicates a difference from a car moving in the opposite direction, and potentially impacts human behavior, if one were to cross the road, for example. This indicates that space holds additional and at times necessary information. The organization of sounds, their movement and location is an organization of associations that in turn form greater structures, communicating more complex concepts. Associations are akin to expectations, which is a major part of musical progression. Playing with expectations through associations gives the ability to enhance or undermine concepts, depending on context. Realistic or abstract sounds both lie on the same continuum of sound vibrations, yet the difference lies purely in their perception, seeing as musical notes are agreed upon reference frequencies.

The practical part of this paper takes on the form of an installation, with the utilization of the interrelational interaction of sound and space in generative sound. The format of the electroacoustical setup is a four-channel format distributed in an imperfect quadrilateral manner, allowing for a margin of adaptation depending on the location of installation. . The loudspeakers are placed in a way that pairs form a single line depending on their combination (diagonal and side lines) thereby rendering them 6 pairs in total. Surround sound in movie theaters played a pillar role in the rise of immersive audio, as it hosted one of the first immersive setups (quad and 5.1 loudspeaker configurations). Theaters were already an established attraction, but mainly for visual presentations. The orientation of the theater is defined, such that the front is where the screen is located. Parallely to theater playback systems, sound designers experimented with different multi-channel playback systems. Karlheinz Stockhausen, Edgar Varese, Pierre Schaeffer, Iannis Xenakis, Chowning and several others pioneered in spatial sound playback, making the assumption that this is the next big step for audio playback after stereophony.

A key point in deconstructing spatial sound structures is to question the nature behind the spatial character of the sounds and their general sequential order. In the next chapter, some of the works of the above-mentioned artists are discussed, and in this thesis, an artistic work will be presented in a four-channel setup and reflected upon, using methods from Artistic Research Methodology (Hannula et al., 2014) .

A Brief History of Spatial Sound

The link between physical sound and otherworldliness comes as unsurprising, as the combination of sound and spirituality is evident throughout history. Any object or space with a strong acoustic resonance has the ability to acquire spiritual meaning. It is our ancient ancestors who first realized the relationship between sound and space. In the cave at Lascaux in France, prehistoric drawings of bulls, painted some thousand years ago, were found in chambers inside the cave where specific resonances peak. Lacking scientific knowledge and having rather little interest in the functionality of human perception, outer stimuli were interpreted by prehistoric humans as spiritual. This led to the belief that these resonances signified greater meaning, a symbol of higher beings or a divine existence. This site-specific art can be called an installation, and in fact, one of the first of its kind. In contrast to the visual domain, which is made of light reflections off of surfaces, sound energy travels in space, as it is for example in a cave, processed by chamber structures and reflective surfaces, and then is dispersed back in an altered manner than its original nature, revealing an inner catalyst. Thus, it was believed that this "spirit" (the spatial sound modification of chambers) has a consciousness or a mind of its own which speaks, namely in the form of acoustics in the physical world. Even in modern day society, it is still common to say that God speaks to us and hears us (Blesser & Salter, 2009), which can interestingly be found cross-culturally. Researchers also studied sonic structures in Great Britain, dating back almost 5000 years, which amplify a resonance of around 110 Hz, the area within which the male human voice is located. More recently, church bells and prayer calls in mosques began to be used to communicate, giving identity to towns or cities. The Acheron Necromancy is a Greek temple of the 8th century with three-meter-thick walls and is reported to have been used to speak with the dead within the tunnels and chambers of the temple. This attribute of resemblance to a higher spiritual being is coupled to the sonic propagation of the information, reaching thousands or millions of listeners. In recent days, the belief that sound has a spiritual essence has not vanished, however with the emergence of science and its developments, more logical views have been increasingly adopted.

Between the great Renaissance and the early Baroque eras, polychoral music began to emerge. This so-called "cori spezzati" ("separated choirs") introduced the separation and allocation of choir groups in different locations in space, singing in alternation and/or successively. This changed the definition of space from being a mere venue hosting an event, to a sound forming enclosure with certain qualities, independent of the music, but dependent on the space. Spatiality, with the effects that it brings to sound, such as delay, reverberation, echo, spectral filtering and dynamics, emerged for the first time as a creative and artistic parameter at the disposal of the composer, who decides not only where the spatial position of the singers/musicians should be, but also their timely cues in respect to their placement. The development of this style of musical performance in return influenced the conception of music writing for spaces.

The relationship of space and sound has been well established throughout history long before electroacoustics or spatial synthesizers were invented. For instance, some buildings built by ancient Greeks offered small rooms for private and quiet conversations between adjacent people as well as other rooms that were bigger, more public and hosted a greater number of persons, where sound reverberated strongly, which did not suit persons seeking to have private conversations. Other Greek venues were open-air and therefore did not allow for much reverberation, a convenience enabling the accommodation of a greater number of people for political, religious, or artistic events. Some of these architectural structures had openings or outlets, which helped in leaking sonic energy outside of the place, in contrast to closed structures like churches. The Roman Empire's change of religion to Christianity gave rise to the transformation of a vast number of these venues into churches. Eventually, and as a display of wealth and privacy, these sound outlets in the architectural structures would then be closed, which led to the reflection of sound within the structures in a more intense manner. Later, worship songs were sung in these venues and the aural architecture in this case contributed, similar to the performances by our early human ancestors, to an atmosphere that promotes spirituality (Blesser & Salter, 2009, p. 90).

The relationship of spatiality and music is evident even in the mid sixteenth century. As previously mentioned, Cori Spezzi is a form of musical alternation where choirs positioned at different locations in cathedrals sounded polyphonically and consecutively. This resulted in situations that affected the composition itself, due to the form of space. For example, the overlapping of sonic energy caused by long reverberation times and the delay of sound in time because of long distances between choirs was in some venues desired for its effect (overlapping notes that formed harmonic chords), while in other venues unsuitable because of its chaotic and undispersed energy leading to decreased clarity. Moretti (2004) writes about how the architectural space of the St. Mark's Basilica imposed limitations on the compositions of music that were designed for two or more choirs:

> Because the choirs are located at some distance from one another, the composer must see to it that each chorus has music that is consonant, that is without dissonance among its parts, and that each has a self-sufficient four-part harmony. Yet when the choirs sound together, their parts must make good harmony without dissonances. Thus composed, each choir has independent music which could be sung separately without offending the ear. Moretti (2004), (Bryant, 1990), Brech & Paland (2015)

Humans first separated sound, space and time as soon as technological advancements allowed it. The development of telephone communications in 1876 and later the radio in the 1890s represents the earliest acousmatic experience. Communicators were not anymore in the same space as the listeners, yet they both still communicated with one another. Sound and space became separate and transportable entities. In 1877, the development of recording devices separated a sound event from the time in which it happened. This displacement has been in emergence ever since (Brech & Paland, 2015) and shall be closer examined in the following.

One of the first electrical instruments, the Telharmonium, dates back to the early 1890s. In the 1920s, the Theremin first came into existence. Lev Sergeevi Termen first drafted a prototype and made use not only of synthesis by overlapping two waveforms, but also of control of the pitch by gestural hand movement in reference to an antenna. This came to be one of the first gestural electronic sound generators (Brech & Paland, 2015, p. 109). Later came the Trautonium between the 1920s and 1930s. The Trautonium was developed by Friedrich Trautwein in cooperation with the Rundfunkversuchsstelle (RVS-Trautonium). It uses glow lamps to function, producing overtones which can be changed using a knob, and has a range of two octaves (Brech & Paland, 2015).

RCA's Mk ll Sound Synthesizer transformed electrical signals to sonic energy, the original concept of modular synthesizers that only appeared in the mid 20th century. Converting electrical signals to physical sound facilitated sound design by using electrical manipulation processes. In 1964, the American engineer and electronic music pioneer Robert Moog developed the “Moog” synthesizer. Moog synthesizers would set the standards for the next decades and are still widely distributed today. While Moogs featured a keyboard which benefits western music styles, in 1967, Don Buchla created a similar yet unique type of playable synthesizer that used playable metallic plates to transmit control voltages depending on contact position and gesture velocity. This type of playability did not divide frequencies into notes, but allowed access to the continuity of the sound spectrum. Thereafter came ARP in 1970 and the E-mu in 1972, all of which are based on the same concept of playable voltages (Görne, 2017, p. 14).

In the early 1920s, it became possible to record sound using monophonic microphones. Emerging as a point in space, monoaural systems served for a decade. In 1931, Alan Blumlein introduced a stereo technique, which was a binaural system, consisting of two eight figure microphones stacked on top of each other at a 90 degree angle. Due to this technique’s valuable spatial image reproduction and correlation to the human hearing, stereo techniques later branched from Blumlein’s concept and spread rapidly in the audio industry. Thereafter, the number of channels would increase proportionally for different productions, where some audio productions in the mid twentieth century reached more than one hundred channels, like the Poème Èlectronique.

Poème Électronique is one of Edgar Varèse’s (1883) most popular works which he presented alongside Iannis Xenakis,who oversaw much of the Philips Pavilion project at the 1985 world exhibition in Brussels. Poème Électronique is an exceptional art-piece, in which a whole Pavilion was constructed by Le Corbusier, drafted and in collaboration with Xenakis, who is a composer and an architect. The pavilion housed 425 loudspeakers organized around the inner circumference of the tent. This was planned on the basis of the harmonic series and modular coordination of three-dimensional continuity (Capanna, 2015). The poem not only had sound that was manually moved on trajectories by Philips engineers, (using switches) conducted by Varése, it was also a multimedial composition encompassing images and lights in accompaniment with the sounds. Edgar Varése coined the term "sound mass". Olivia Mattis says a sound mass is: "a composite sound that has a recognisable timbre, rhythmic profile and articulation, but whose components can be altered one by one to highlight different aspects of the sonic gesture" (Harley, 1994). Varése argued that another strong form of sound separation, as spatial separation and dynamics are, is timbral quality. He also had the dream of having his sounds projected into space, which he realized in the Philips Pavilion. He spoke of forms and flocks of "intelligent" sound objects moving in space which have a distinct behavior, and of geometry of sound. He added that space is literally a psychological association or symbol. Varése assumed that because music does not happen without the vibration of air particles between the instrument and the listener's ears, music must exist in sound, and therefore is of physical qualities that can be objectified. He foretold how in the future, new means would make transmuting planes and sound masses possible. Music would also take on a “flow” state and time would cease to be static, where sounds have different speeds and directions. The “Poème Électronique”, the performative installation, is not reconstructable, as after the world exhibition, the pavilion was taken down.

Pierre Schaeffer, born 1910, first hinted at the change in instruments brought by technology in the nineteenth and twentieth century. Also founder of *Groupe de Recherche de Musique Concrète,* using technology he developed the concept of “Musique Concrète”. One of the most renowned electronic music composers of his time, he contributed extensively to the scenery. He used the now technologically primitive turntables to playback multiple pre-recorded sounds and manipulate their speeds, deform them, and alter their sequence. He used the studio as an instrument, so that its parts contributed to the end product as a whole. By choosing the sequence of sounds, as notes and the distance between them would be chosen, a narrative started to develop. He also brought spatiality into sound, as Brech & Paland (2015) write: "Technology can harness our desire to ›play‹ spaces with embodied movement and gesture". Musique Concrète, as he defines it, first used natural and fabricated sounds, whether recorded or synthesized, and manipulating them similarly to cinematic audio techniques. Secondly, the projection of the sounds into a three-dimensional space creates depth. Pierre Schaeffer was concerned with the question of whether it is possible to create orderly structured material out of the richness of nature and of synthetic sounds, a symphony of noise as he would call it. He not only helped in the genesis of modern electronic music, but also aided in creating technologies for it, in cooperation with the engineer Jacques Poullin. One of his techniques was to sample recorded sounds to pieces and to layer them above each other in order to achieve a new sonic quality. Led by this technique of creating sound, Schaeffer patented a device with three sound sources, with a routing switch enabling the overlapping of the sources. Being concerned with making music as much as noise, he developed and patented another three-source playback device based on his older device, which enables the playback of all three channels simultaneously at different speeds. Later, the well known *Phonogéne,* a precise variable-speed (varispeed) recorder and playback device would also be added to his list of patents. Other performative devices were also drafted as a need for a musical presentation, such as wired controlled potentiometers with a foot pedal to control low frequency amplification. Thereafter, in the 1950s, Pierre Schaeffer concluded an important milestone in spatial sound, as Adrean Willaert had done in his time. Pierre realized that his noise or sounds are not only foreign to the human ear, but that they should also not come from a single mono source, which was very common at the time. Rather, as the noises were new, so should the spaces in which they find themselves be. He argued that monophonic and even stereophonic sources are limiting the possibilities of the spatial exploitation of sound, and needed to be liberated. Two types of spatialisation techniques were thus identified by him. The first, a static composition of sounds placed in space, and the second, moving sound sources (Brech & Paland, 2015, p. 129). Limitation of left and right channels started to vanish as rear and over-head loudspeakers were added, giving depth or three-dimensionality. Indeed, this was the beginning of immersion in electroacoustics.

Henry Brandt, born in Montreal, Canada in 1913 was a leading force in multi-textural and polyphonic music. He assumed that all music is spatial music and worked with multi-percussion ensembles in the 1950s. After realizing that, although his musicians were playing in time quite accurately, problems arose in polyphonic percussive pieces with a large number of percussionists (more than 60) over larger spaces (like in the San Giorgio Maggiore). As he listened to his own music being played with this many musicians, he himself was quite confused. After learning about Giovanni Gabrieli and the Venetian School, the Cori Spezzati, he realized that separating the percussionists spatially would lead him closer to the polyphony he was striving for. He also realized that this had implications on the musical composition. Henry broke out of the stiff bars and measures prescribed by music theory and became more flexible in writing for percussion, in a manner that was not rhythmically rigid. In 1952, Teo Macero, a student of Brant, composed and performed Jazz music for five different bands placed at five different angles. Brant also experimented with the correlation of pitch and height, by trying various positions for musicians and locating lower-frequency instruments at a low horizontal level and high-frequency instruments on concert hall balconies, as well as experimenting with the effect of pitch on vertical locality, saying:

> The actual pitch need not to be high or low, as the case may be; if the register in which the instrument plays is proportionately acute or deep, this will substitute very well for absolute height or depth in pitch, and the instrument (or voice) may be situated in, respectively, high or low positions accordingly. (Brech & Paland, 2015, p. 94)

Thereafter he then turned to experimenting with moving sound sources by composing pieces for musicians that changed their places during the performance.

Iannis Xenakis, who also composed the preceding piece at the Philips Pavilion in the Brussels Expo ‘58, as mentioned previously, was born in 1922. In France, he met a number of people and specifically those in the Le Corbusier office. He focused on numbers, trying to bind space and sound in his works. Noteworthy about his works as a mathematician, is that he embedded stochastic concepts in music, which briefly defined, is the use of probability as a determinant. Next to this technique, he also touched on other theories like play theory and set theory, to name a few. The concept of tamed randomness and probability is predominant in generative sound (Harley, 1994). Although the music of Xenakis was not generative but pre-composed, stochastic rules were applied in the compositional phase.

Pierre Boulez was born in 1925. He was a French composer, music theorist and director. He crossed roads with Pierre Schaeffer as a student in Paris then went on to be one of the great influencers of electroacoustic music. He also worked with Stockhausen and later participated in academic teachings in Darmstadt, Germany. He specified different attributes to sound that make a sound unique, like frequency, loudness, duration, timbre, and spatial function. Moreover, he stated that movement can be applied to all of these faculties, and not only to spatial location. Remarkable about Boulez's suggestions that differ in nature from those of other composers like Stockhausen, is that he considered space not an ornament or a feature of sound, but rather a quality of probability (index of distribution), and its core not being in its aesthetic quality, but in its function. Therefore, Boulez was not a fan of using space to invoke mystical effects. In contrast to Varése, he was less concerned with the geometrical shapes of sound movement and displacement, as all geometrical shapes can be achieved within a circle. It is rather the distribution of these sounds according to a set of rules and probability, as well as the speed in which they behave, that interested him. Separation happens through the pitch, dynamics, timbre of sounds and so on, regardless of their spatial space. He adds that movements have to not only affect the sounds themselves, but also their microstructure. Movements become interesting when they have a "brownian" nature referring to the chemical movement of particles that can not escape bumping into each other, enhancing or influencing each other. Time also plays an important role, as our perspective of linearity of the order of events is static, sounds have the ability to be sounding either linearly or simultaneously. *Conjunct* and *disjunct* intervals, he explains, have to do with the overlapping and separation of the sounds in time, respectively. He also gives importance to homogeneity of the sounds that emerge, pointing out that homogeneous sounds share a connection or perhaps may be perceived as a single sound source, regardless of their location. Spatial symmetry of sounds can be achieved, regardless of the positioning of sounds, and whether they are homogeneous (of same or similar timbral quality for instance, rendering regular symmetry) or nonhomogeneous (irregular symmetry). So in brief, the definition of spatial Sound to Pierre Boulez was less similar to that of the likes of Stockhausen or Varése in some aspects, in that Boulez leans toward discarding the motion of sounds, but highlights the emergence of sounds at different locations, with the emphasis on sound qualities like timbre and pitch. Similar sound qualities sounding from different locations and overlapping simulates sound movement. It was of less ornamental concern to involve motion in sound, but rather a function of clarification. Spatial location helps the ear discern single sound elements from another, increasing clarity and comprehensibility (Harley, 1994).

Karlheinz Stockhausen, born 1928 in Germany, experimented with sound synthesis and pioneered in spatial sound. A lot of his work included sound that was recorded on a one- or two-channel tape, manipulated, and then distributed in space (Stockhausen, 1981). He developed devices in order to help him translate his ideas much like Schaeffer did, e.g. the rotating table, as well as other devices in collaboration with WDR Electronic Music Studio. He has a number of prominent works like Gesang der Jünglinge, Kontakte, Momente, Hymnen, Sirius, Helikopter-Streichquartett, to name a few. In a lecture in 1972 in Oxford, Stockhausen spoke of how he used the basic knowledge of the first half of the nineteenth century and expanded on it. Traditional instruments were so well established that the only few electronic instruments available at the time that he started experimenting with around the mid 1950s, could only imitate these traditional instruments. Stockhausen found interest in developing new sounds. He also brought forth the *four criteria of electronic music* *composition*, which are also criteria for spatial sound composition. The criteria are as follows: *time structuring, sound splitting, multi-layered spatial composition, tone and noise quality*. In time structuring, the importance of being able to speed up or slow down a sound continuously is key, as Pierre Schaeffer had done. An example of time structuring is speeding up a Beethoven symphony while preserving its pitch so that it becomes only one second long. The piece would now be a small tone that has a particular exclusive textural quality and is derived from the spectral and temporal structure of the original piece. Similarly, Stockhausen talks about stretching sound that has a fairly short duration (1 second) long enough (about 20 minutes) so that it becomes a musical piece of its own, magnifying its acoustical spatio-temporal phenomena, which reveals the micro-acoustical structure recorded at that time. Working with the criterion of *tone and noise quality*, one can exploit the possibility of how rhythm is perceived as a pitch when it is looped and sped up to a certain frequency. Rhythm hence becomes pitch. Periodicity and inner subdivisions, or partials, determine the timbre. The change of periodicity changes the timbre systematically. Sounds of a distinct texture have a definite or cyclical periodicity. Noise, on the other hand being aperiodic, is a random or chaotic change of the periodicity. Sound (and music) is a construct of time continuum, where pitch, harmony, timbre, rhythm, melody, meter, periods, phrases, sections, and larger musical entities are dependent on this construct. Stockhausen also identifies the spatial language of sound as flawed and lacking in original vocabulary, since high frequency sounds are called *high-pitched* where perhaps the description fast-pitched suits the concept better. This is an indication of the attribution of space to sound. In *sound splitting*, he indicates that a sound could be put together, but may not be stationary. A non-stationary sound changes over the course of its time, like the word "now" as an example. Karlheinz Stockhausen speaks of composition and decomposition of sound. Layering multiple sounds or parts together to create a whole sound is *composition,* whereas *decomposition* is the taking-apart of a sound and listening to its parts alone. Interestingly, the profession of composing encompasses the art of putting instruments, rhythms and melodies together, be it in music or in organized sound, to make a greater whole. Components, when subtracted from or added to the whole, change the characteristics and quality of the whole. Furthemore, according to Stockhausen, the use of space is changing. The perspective in which humans have found themselves is a fixed perspective. He emphasizes composition, movement, and the speed of sound in space, which are becoming as important in the composition as notes and melodies. He adds that depth is an important parameter when we talk about a *multi-layered spatial composition*. It is not only possible to move the sound in any direction around the listener, it is also possible to move it closer or further away from them. From his works at the time of the lecture, he has managed to accomplish superimposing up to six layers of sound acoustically, but has found it hard to increase the number of layers to more than six, since it becomes harder to follow between the single sound effects. He goes on to explain that in the work of "Kontakte", he establishes a foreground or a wall of sound which is heavily dense and hard to penetrate. Afterwards, the wall of sound is stripped down layer by layer, taking away its parts, exposing the underlying, unconscious layers of the sound to our attention. With every layer faded out, a new unnoticed sound is revealed. In comparison to the two dimensional melody, harmonic and rhythmic composition, this adds an intrinsic parameter to control. Stockhausen (1972) says: "This happens everywhere in nature. It's very important to hear if a car is far away or if the car is two feet away, [depending on which] I will behave differently", indicating the behavioral impact spatial proximity of objects has on humans. There is also a difference between loudness and distance, as loudness is independent of distance, which has been made easily manipulable with the help of electroacoustics. He continues to explain the present lack of experience in categorizing new, acousmatic sounds, due to spatial and proximity cues becoming harder to grasp with unknown sounds. He names three features of a sound that determine its proximity: *intensity* in an absolute sense, *degree of distortion*, and *dynamics*. Here, Stockhausen indicates that in having control over these three categories, proximity can be simulated. He elaborates that for sound to be perceived correctly, there should be a reference for the sound, as well as crossmodal congruence. Sound events are a result of objects that the consciousness would like to have identified. Moreover, there is more to noise than it being random or chaotic modulations, namely, *noise quality*. Noise quality can be determined by density, or the concept of how much energy a sound has, and in which spectral segment. Stockhausen clarifies that in electronic compositions, care should be taken concerning the noise to note ratio. Noisy tones for instance are far more destructive than clear, sine wave-like tones, which might render a composition too noisy or disturbing. The difference between microtones of sine waves for instance is clearly perceivable, whereas overtones and notes make this difference harder to perceive (Stockhausen, 1972b).

Parallel to the concept of using modular synthesis to spatialize sound, the Halaphon, a spatialization tool, was developed between the 1960s and 1980s by Hans Peter Haller and engineer Peter Lawo. As Luigi Nono worked at Freiburg Experimental Studio, he created a spatial music piece called “Prometeo, tragedia dell’ascolto”, using the Halaphon. It is a device used for live presentations and was updated time and time again, with even digital technology being incorporated in it by the end of the 1980s. A somewhat more advanced tool than the techniques used by Stockhausen, the Halaphon used amplitude panning to position or move sounds across a loudspeaker setup. This panning technique generates a phantom image, which is the illusion of a sound source being positioned or moved between two or more loudspeakers, by manipulating the volume ratio of the loudspeakers. In a nutshell, the device controls signal volume between at least two loudspeakers in order to make the illusion of a moving sound object. Using voltage envelopes triggered by gates and fed into amplifiers, the signals were amplified, and, depending on the polarity of the signal, either increased or decreased the volume of the channel or loudspeaker. The Halaphon also had a preset manager and a large amount of preset saving capabilities, as new Halaphon versions were developed. The presets contained the point of departure of the sound, as well as the form and duration of the voltage envelopes. It was possible to manipulate the trajectories during performances which made it suitable for live situations. Halophon was used with other gear such as analog devices, as well as reverberators and other signal effects (Brech & Paland, 2015, pp. 194–204).

In as early as 1967, John Chowning was already working on digital frequency modulation synthesis as well as its spatialization. Chowning, who was born in the United States in 1934, played a tremendous role in electronic sound, as he brought forth several developments into the discipline. In almost all modular synthesizer oscillators, there exists an input for frequency modulation. FM synthesis started to pop up in electronic music instruments and playable interfaces, becoming available for musicians and sound designers to experiment with. It was also used to synthesize known sounds, like the human voice or bells. Chowing also pioneered in the four-channel format and in bringing space into artistic context. He composed several pieces for a four-channel system (quadraphonic). In the field of psychoacoustics in enclosed space, Chowning based his theory on two variables that determine position of a sound source: *angle*, which is determined by the difference in the time of arrival of sound at both ears, and *distance* of the listener to the sound source depending on both sound pressure level, including the amplitude of high frequencies (which are also affected by the head shadow), and the ratio of direct to reverberant signal

(Chowning, 1977). He assumes that the premises within which the four-channel setup is, is called a *listener sound space,* while the area outside of it is an *illusory sound space* (See Figure 1.)

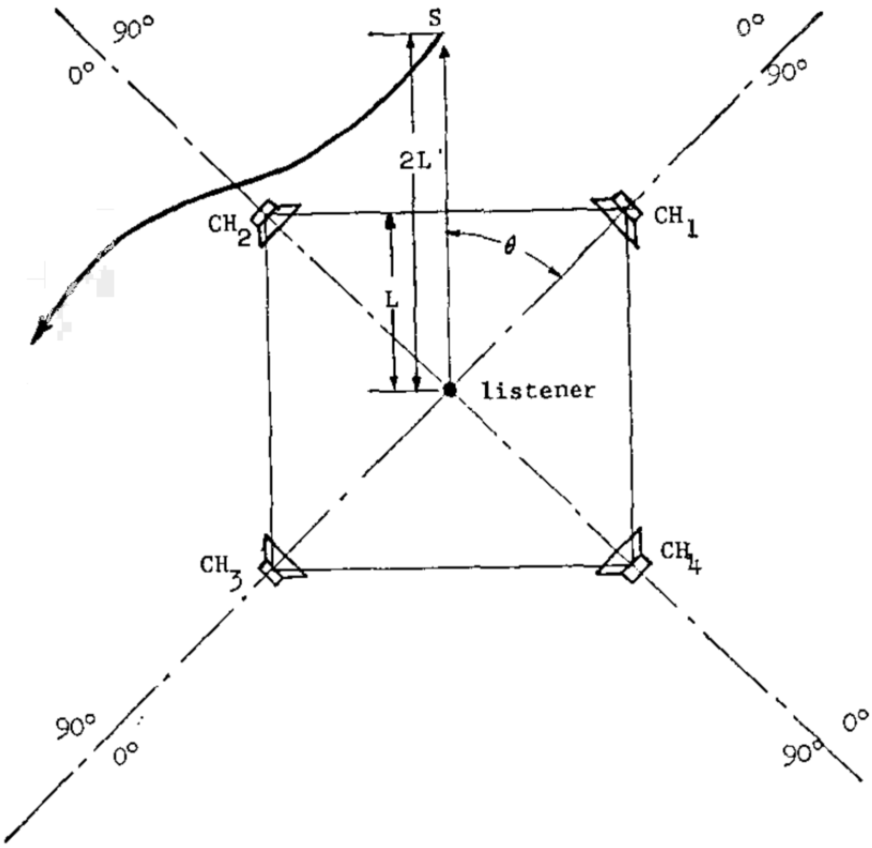

Figure 1. Configuration of loudspeakers defining illusory and listener space. Chowning, J. M. (1977)

Sounds were spatialized in that the intensity of a sound is divided by two and distributed to two loudspeaker pairs. In simulating distance, the focus was to differentiate between close and far sounds with varying loudness, like a soft, near sound and a loud, far sound, with both sounds having the same intensity. Similar to Stockhausen, Chowning claimed that spectral frequency shifts, such as Doppler's effect (up-shift of frequencies of objects moving towards the listener and down-shift of frequencies when moving away from the listener), or the compilation of resonances that make up reverberation, impacts perceived loudness independent of intensity. Vibrato, for instance, induces instability and makes a sound appear louder. The ratio between direct and reflected sound indicates distance in reverberation, whereby early reflections influence loudness. These attributes were considered in Chowning's spatialization criteria and included in a digital program that allowed for angular and radial velocity computation in the early 1970s, with an interface for plotting points of sound trajectories. Chowning also addressed sound composition and the timbre of sounds. He put forth that, in pursuit of natural timbres, like the ones found in nature, *micro-modulations* shall be equally injected into the waveform generator, for instance into the frequency partials and/or their amplitude. This causes *perceptual fusion,* the phenomenon in which multiple alterations of sound are perceived as a tone. Nevertheless, injecting non-equal modulations to different clusters of partials leads to the ability to separate between different tonal elements, which he called *source-segregation*.

Dieter Doepfer (1997) was a then aspiring physicist and is now a modular synthesizer producer. One of his first modules was a voltage controlled phaser for the Elektor's do-it-yourself Formant synthesizer. Later, working at a hospital as a physicist, he had the time to develop some modules of his own and started selling them. In the early 1980s, he developed a polyphonic system that had great reception. While working on a small scale, Dieter Doepfer also had an eye for the potential of digital technology and seeked programming support. As he tried to experiment with digital technology, it was under-appreciated and had not received enough credit. In 1995, Dieter Doepfer established the Eurorack standard for his modules and developed a full system for it, after having previously found success building keyboards. Now, with hundreds of manufacturers and over a thousand modules on the market, Eurorack was not very popular until after 2010. Doepfer Musikelektronik GmbH set the standard for Eurorack modular synthesizers (Ableton et al., 2015).

Project Concept

There are already several immersive systems with variable channel counts being researched in the audio industry, and their refinement is an ongoing process. There are, however, fewer sound-design tools or musical instruments with integrated control over the spatiality of sound as a design attribute. In an attempt to combine and integrate the qualities, techniques, and features mentioned by Boulez, Varése, Schaeffer, Stockhausen, Chowning, and others, and to study the interdependence and reciprocal effect of space on sound, a spatial modular synthesizer was drafted. A generative installation is realized as the practical part of this project. Based on four channels, every channel corresponds to a loudspeaker forming a quadrilateral shape, some of which are found in the diagram in Figure 2. There is a possibility of six phantom images in total (all the lines between all the loudspeaker pairs).

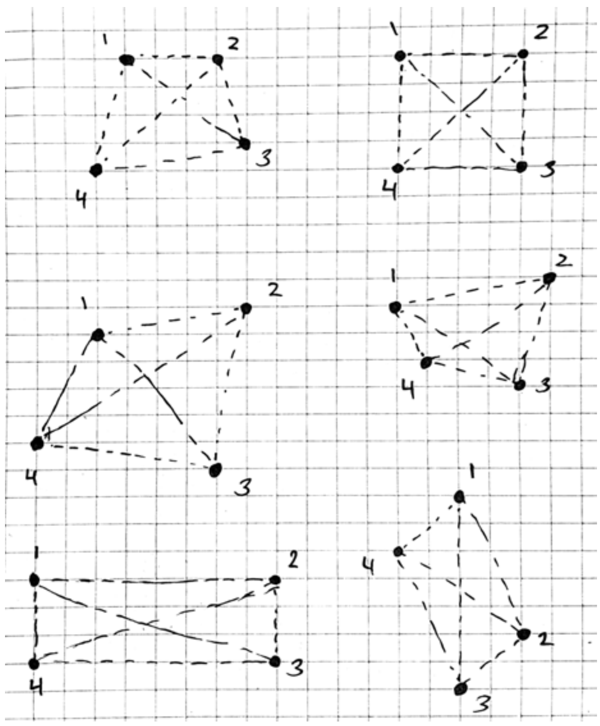

Figure. 2. A sketch of loudspeaker placement setups trialed

Quadrilateral includes, but is not limited to, quadraphonic configuration. The distances between the loudspeakers and their respective angles were set up in a way that somewhat coherent spatial images can be composed (in that loudspeaker pairs were not too close or too far from each other). All loudspeakers were at a distance of between 1 meter and not more than 30 meters apart from each other. The configuration will be called a four-channel or quadrilateral loudspeaker setup throughout this thesis. There is no particular orientation of the setup. Front, back, and sides are not defined and can be attributed to any loudspeaker or loudspeaker pair(s). This is subject to the listener's position and movement. Although spatial sound can definitely be achieved with as few as one or two channels, like the binaural experience or using the crosstalk cancellation method with loudspeakers, the four-channel configuration seemed most feasible and suited to the project's realization. This is because the research's intention is not to simulate binaural experience. While a three-channel format provides the possibility of three-dimensional sound, only a single triangle (lines between loudspeaker pairs) can be achieved. With the addition of one extra loudspeaker, the number of lines between loudspeaker pairs doubles (for example: triangle 1 is constituted of channels 1, 2, and 3, with a diagonal line between 1 and 3. Triangle 2 is constituted of channels 2, 3, and 4, where a diagonal line is found between 2 and 4, and so on). For more clarification, see figure 3.

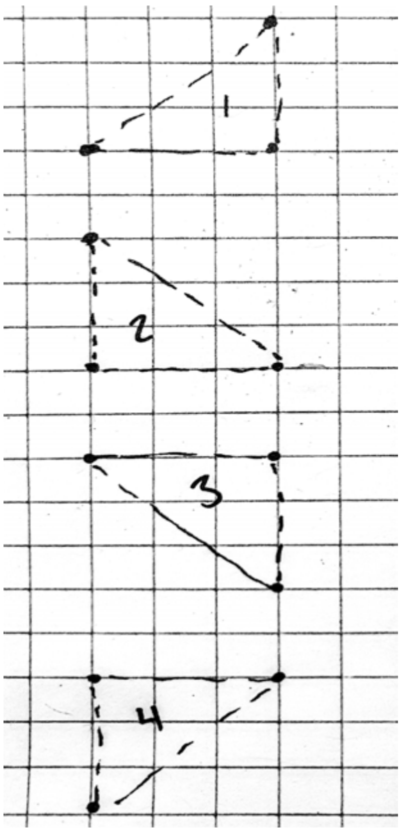

Figure 3. A sketch of possible triangle formulations in a four-channel format.

The importance of the lines between loudspeakers lies in the illusion of phantom images, as well as other psychoacoustic effects which will be discussed later.

The instrument drafted and assembled for this project not only touches on the concept of amplitude panning of the Halaphon, but also expands its spatial capabilities, integrating spatial synthesizers like reverberation and delay, manipulating spectral content and cues, as well as the interaction of the aforementioned attributes with the sound and its behavior.

The process of patching a new instrument every time contributes to installations in particular, as sounds can hardly be replicated, as well as to the corresponding trajectories and to spatialization. This becomes a tool which enables the development of techniques that integrate space into sound, and vice versa. Both can be created simultaneously in a modular synthesizer, resulting in an interaction of sound and space during the conception, where sound and space determine each other's qualities. Questions arise such as, how does spatial character of a sound demand that it sounds and develops? How would a sound of certain qualities behave in space? The convenience and freedom of modulating sound and space using common modulation sources induces intriguing effects. With that kept in mind, it was attempted to mimic, resemble, or project internal sonic modulations and fluctuations (like pitch or timbre) spatially, bringing the partial, qualitative sonic movements outwards and projecting them physically.

Theory

Perception of Hearing and Space

The Psychology of Spatial Perception

> If I had to choose between being a person who knows what beauty is, or a person who knows what sensing (perception) is, I would choose the latter, for when and if this knowledge becomes available, although I'm afraid this is merely imaginable, it would reveal all secrets of art. (Valéry, 1937, p. 2)

As established, perception is a fundamental aspect of art. There are five stages to the perception process: Stimulation, Organisation, Interpretation-Evaluation, Memory and Recall (Devito, 2009). The brain finds itself in an active course of this process. In order to perceive, humans have developed a range of senses consisting of vision, hearing, tasting, smelling, and feeling, which suffice as the main stimulus inputs. In this artistic research the main interest is the sense of hearing. References and analogies to other abilities of perception will be made. Beginning with the first stage of perception, stimulation is an event or phenomenon that happens in the "outer" world. This stage encompasses attention, defined by (James, 1950) as: "the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, and consciousness are of its essence". Attention can be selective. For instance, E. Colin (1953) made experiments on speech recognition and subsequently, the now well known cocktail party effect was discovered (Bronkhorst 2000), which found that people tend to hear and react to their names being called, even if other abundant stimuli is present. This is called selective attention, and has been proposed by Dalton & Fraenkel (2012). Another example of this is simply the act of one focusing on a task at hand, while tuning out other stimuli. In the organization stage, it is creating a scheme, categorizing the information so that they are easier to recall later. This is exactly why humans tend to recognize patterns more efficiently than scattered or random information (Metz-Göckel, 2016). Interpretation-Evaluation is where the meaning of the stimuli is allocated. This is subjective in its form and describes the inner reflections on the outerworld. For instance, one might like ice-cream, but finds a particular flavor undesirable, hence personal preference defining an object. In the memory and recall stage, memory serves as the tablet for cultural and learned metaphors of perception, proceeding the stages described above, and stores patterns and schemas for later recall, much like a memory card. Recall is the last stage where the saved information can be revisited from the memory.

> (...) taking any particular detail, a person who remembers can set it into relation with (an)other detail, stating its setting uses the organism, so to speak, to produce a differential reaction; in remembering, the subject uses the setting, or scheme, or pattern, and builds up its characteristics afresh to aid whatever response the needs of the moment may demand. (Metz-Göckel, 2016, p. 172)

Perception is the simultaneous happening of its three parts, according to Husserl, namely *retention*, *primary impression*, and *protention*. This categorization also being the partitioning of time, retention takes on the form of past tense and is the incremental addition of moments just passed (in the memory). *Primary impression* defines the ongoing momentary awareness and happening, manifested in the now. Lastly, *protention* is the anticipation of the future happening. There have also been categorizations of memory, as recollection differs, on the chronic level. *Echoic memory* takes on the continuous stance, as it is ongoing and does not halt. *Short term memory* stretches only for a couple of seconds, while *long term memory* goes beyond that. A quantal model for auditory objects has also been suggested, transforming a stream of events or awareness into chunks of memories, where a two-way temporality of memory is assumed (the recollection of memories at different times in a non-linear fashion) (Bader, 2013).

Senses are in many cases independent from one another. That means, we perceive an impulse in independent mediums until the brain is reached, where processing in independent brain-modules happens. While clapping for example, we perceive the impact of the hand with touch, vision, and hearing. Though these events are processed in the brain individually, they are perceived as a single event. This fusion helps to understand the environment, yet the ability to perceive individual signals still stands, and helps in differentiating clapping from another impulse like thunder, for example, which has fewer redundancy. This knowledge that is gained through the senses contains location cues that are redundant, as they are concurrently available (Waller & Nadel, 2013, p. 85). Therefore, having one medium of sensing lowers redundancy, in contrast to multiple mediums. Waller & Nadel (2013) state that "Multisensory integration is more likely when signals arise from approximately the same spatial location and are temporally coincident". As indicated before, prior encounters and experiences acquired through the senses also shape and anticipate the outcome of our perception. It is not a new perception when clapping is heard, seen and felt, but is also compared to our prior experiences of clapping and confirms that a set of conditions defined by clapping is fulfilled.

Listening is the perception through and act of the auditory environment. When listening takes place, a complex process in converting physical stimulus into electrical signals takes place, which is then followed by interpretation and meaning allocation, with a possible impact on one's mental and emotional state. From a single sound, it is possible to gain relevant information about an event that caused it, as in what the objects of participation in the interaction are (what they are made of and other features like location), solely from spectral information of the sonic energy and the changes it undergoes. Concerning location, humans are capable of decoding the location of the object, as well as the location in which the event happened, using sound. Emmerson (1998) claims:

> We know (or believe we know) the sources of the sounds we hear in a recorded performance. But in music which exploits the acousmatic condition, such supposed one-to-one correspondences are challenged and this 'search mode' may be brought back into play. (pp. 135 - 140)

More will be elaborated on this topic in the coming chapters. The sound source, which is normally an excitation resulting from an interaction of two objects, illuminates the space it finds itself in due to the physicality of sound travel (reflections, absorptions, and other processes). It is mainly a connection, a sensation, of the outer world to the individual inner experience of it. Blesser & Salter (2009) reported that "raw sensation is predominantly a biological property". Although language decoding requires vocabulary insight in the language (consistency across a single culture), listening to an understood foreign language does not immediately grant meaning, as information is also housed in the intonation and context of the situation, hence carrying symbolic meta-information within, independent of the word meaning. Associations also play a role at the personal level of the receiver. In some cases, listening to sounds may not communicate anything new. For instance, sitting in a classroom, where outside the class window a jackhammer is in operation, the mind of most students sitting in that classroom eventually filters out this constant sound due to its relevance being low. Attentiveness though, can bring the jackhammer into a focal point, making it more relevant.

Concerning the physicality of sound, vibrations in a given medium reach the outer ear, which vibrate the eardrum and subsequently the auditory bones of the middle ear. The stimulus is then conducted into the inner ear and the resulting signal is then turned into a (electrical) nerve signal that reaches the brain (see figure 4 and 5). Vibrations are physical and materialistic, therefore have a measurable frequency, duration, direction, and pressure level. First a vibration reaches the inner ear through the complex mechanical structure of the outer and inner ear, finally reaching the brain as a stimulus. This is the moment where it is perceivable. The perception of the stimulus is an independent entity of its physical counterpart, thus making it unmeasurable with the tools used for physical measurement. For instance, the use of the metric system to measure a thought is inappropriate. It is also important to mention that vibrations are a result of an impulse, and not the impulse itself. The impulse of the interacting objects and the objects themselves also differs in terms of measurement, than the resulting vibrations (Dickreiter, 1982, p. 108).

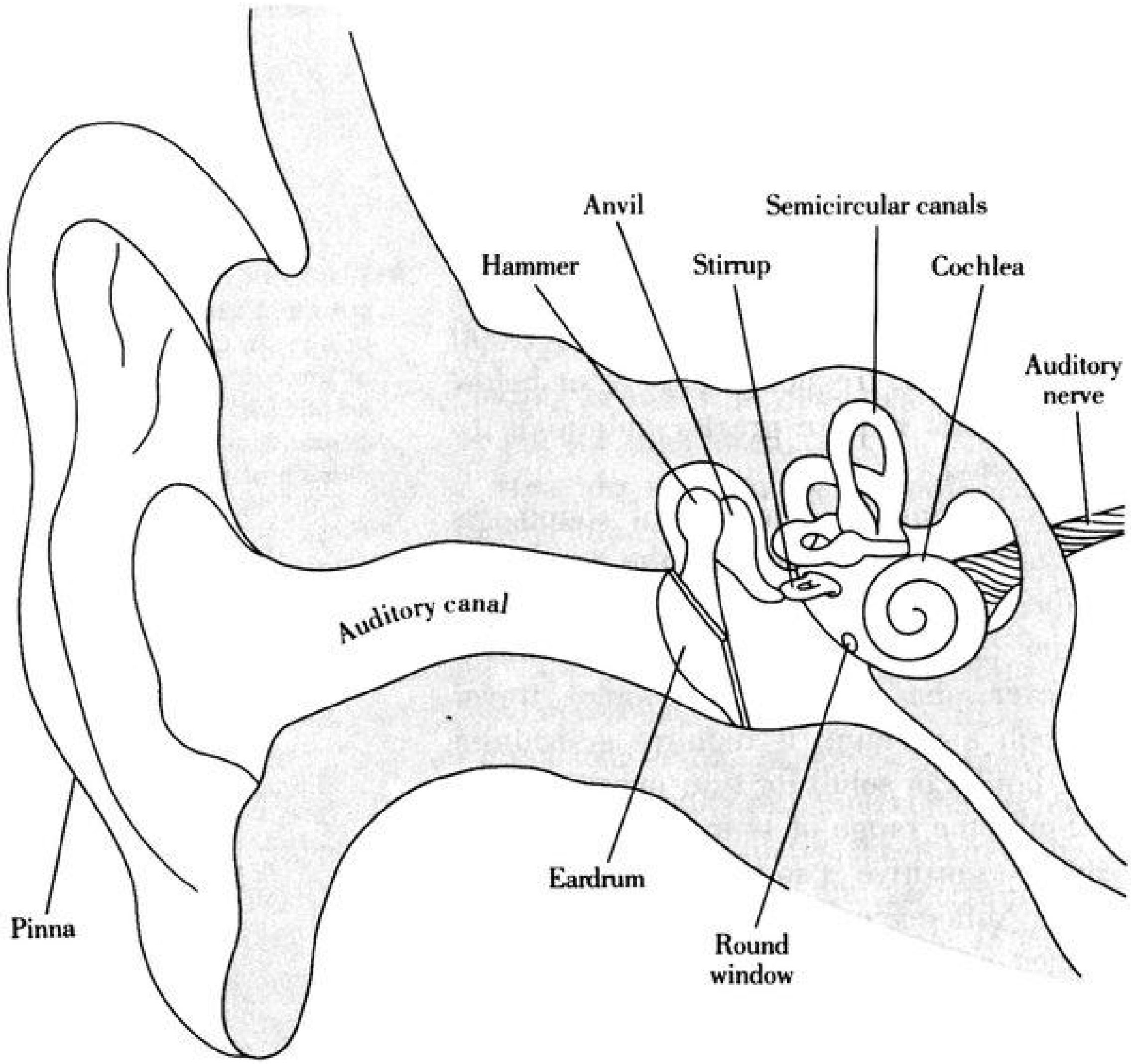

Figure 4. Human Ear Hearing System. Bhandari, P. (2017)

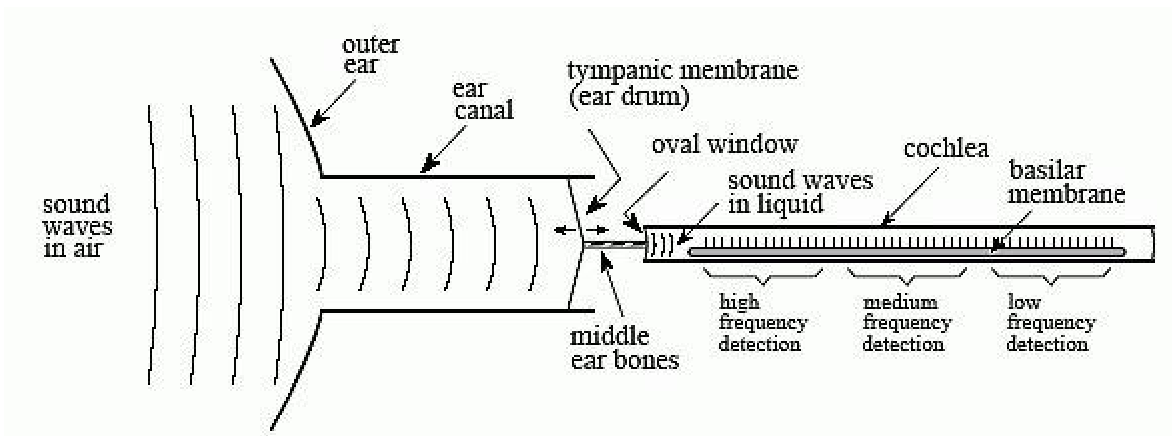

Figure 5. Human Auditory System. Bhandari, P. (2017)

In determining aural location, Interaural Time Difference (ITD) and Interaural Intensity Difference (IID or ILD for Interaural Level Difference) play an important role, but share importance with other aspects. Spectral cues for example, are also necessary in determining location of sound in an environment, as ITDs and IIDs have proven to be erroneous when investigated alone, especially in the median plane, as well as front-back, and distance localizations. IIDs generally function superiorly to ITDs in the low frequency range, while IIDs function inferiorly to ITDs in the higher frequency range. It is to be noted that monoaural information is acquired when modulations are caused by organic movements (such as those of the head) which in turn help acquire differential changes in the sound spectrum and arrival times, thus giving unique cues to every individual. HRTF, or Head Related Transfer Function is the measurement of how sound behaves when it reaches the ear, and focuses on individual differences as well as head movement. HRTFs are, like fingerprints, unique to every human being (Waller & Nadel, 2013, p. 83). It is important to note that humans have spatial reference systems. In order to orientate, locate, or even apply the principle of mental spatial exploitation, reference systems must exist. For instance, the habitat of an individual, which is often allocentric, is defined by the country or city which they inhabit, as well as the house number, street name, and postal code. These serve as reference points for other moving objects or beings (Waller & Nadel, 2013, p. 179).

Spatial Cognition is the ability to perceive the environment around us and process objects and their respective positions. This encompasses size, shape, form, scale and its relation to other objects, as well as distance and orientation. Indeed, spatial cognition also encompasses language (Waller & Nadel, 2013, p. 3). It has been suggested that humans best perceive the world with their body as a reference. Horizontal orientation was best referenced to left and right, while head and feet were used for the vertical orientation. This specifies the link between human bodies and the outer world, referred to as a spatial framework (Waller & Nadel, 2013, p. 235). Most actions and a multitude of mental processes have spatial ramifications. Some mental processes involve spatial cognition without physical maneuvering, like remembering where one placed their shoes or describing where a friend lives. Spatial cognition is the study of the mental processes that happen when space is perceived. During perception, an active process, a spatial representation of the environment or object is mapped and saved to the Hippocampus. It is imperative to distinguish between *egocentric* and *allocentric* representations, in that *egocentric* representation defines the body axes and mapping from one's perspective, whereas *allocentric* representation is separate from the perceiver (Waller & Nadel, 2013, p. 36). The point of reference here is important, as human beings possess the egocentric perspective. Nonetheless, humans develop allocentric maps and can have allocentric thoughts and perspectives. As both perspectival points of view are connected, it is necessary to start with an egocentric angle in order to produce an allocentric standpoint. Egocentrically, the environment can be defined as in reach versus not in reach (Tversky, 2003). The environment supersedes the internal processes and imposes psychological structuring, internalizing the setting and defining the relationship of an organism to its surrounding. When speaking of the outer world, scales are a useful tool. Scales can help reference and facilitate geographic spaces. They are also relative and can be continuous. Scales are not separate from their environment but rather define the relationship between an entity and its environment. Waller & Nadel (2013) state: "small-scale spaces are not necessarily small in physical extent but consist of those environments that require relatively small amounts of action to apprehend, perhaps only eye and head movements". A large scale object or surrounding like a city can be also psychologically regarded as a small scale environment, for looking at it on a map only requires eye movement. Small scale objects, ironically regardless of their size, are objects which we can look upon and not from within. As to medium scale environments, they are typically settings which do not require vehicles to commute within, for example a city park, which enables organisms to move within and also interact with it. Large scale environments necessitate the use of vehicles in order to gain information about them (Waller & Nadel, 2013, p. 4). More information about scale/size distinction are to be found for example in Tversky's (2003) writings. Research by Hegarty & Waller (2005) has found that there is a rather weak link between small scale and large scale environments, which leads to the assumption that different mental processes underlie both scales. This means that the difference also possibly has an effect on the emotional impression of the perception of space. One must distinguish between two kinds of spatial mental processes, namely online and offline processes. These spatial processes differ in that the first is always active and involves the direct perception of the environment while the latter ciphers, remembers, and recalls spatial information. Here, and like in most psychological studies and theories, nature versus nurture plays a role in the discussion (Waller & Nadel, 2013, p. 6). Distance and Direction belong to spatial information. Both can be observed from an egocentric or allocentric point of view. Perception of elevation has yielded much less research than horizontal. It is worth mentioning that spatial information not only gives static clues as to where what is, but may rather imply active happenings and change. For instance, "the car is at the gas station" also implies that the car is at the gas station to be filled up with gas, as normally, that is the reason cars stop at a gas station. As it was not given what the car at the gas station is doing, nor was it confirmed that it is there to be filled up with gas, this logical structuring of the perceived information plays a role in deciphering the bigger picture. Another more concrete example is "the finger is on the button". This almost always implies that the button is or will be pressed. These motorized descriptions are borrowed from our prior knowledge and experience of the world and the attributed functional meanings to objects, and likewise spatial relations. This promotes the belief that the spatial dimensions of objects are not solely the main influence on the association to the world, but also the use of these objects. Lastly, there are six functions for the geographical perception of humans: routing through navigation, learning and applying spatial insight from doing, spatial symbolism, language, mental imagery and logic, and lastly position and allocation. Interesting for this research is symbolism, which is important to acousmatic sound. As environments are grasped through direct learning, finding out where a city is in South America does not need direct experience or learning of the location, but rather a map to pinpoint it (reading and understanding a map is, in this case, a prerequisite). Maps are a symbolic system that humans have developed in order to get an allocentric point of view of the world. It is a way to view information and its representations. Additionally, there are various mappings to depict space, depending on the purpose of the mapping. Some maps may convey information and metric specifications about horizontal or vertical dimensions. Others may specify both elevation and width.

Another example is graphs, as Tversky (1997) reported, which are another way to use space metaphorically to represent information. Symbols are indeed crossmodal in the human senses. For instance, as a map represents a geographical location, so would a distinct reverberation represent a specific church. The same way maps differ in context and kind, so too does reverberation embody different information and convey different sets of informative qualities (Waller & Nadel, 2013, p. 252).

Multiple sound sources that are moving in a single direction tend to be perceived as a cluster. This has been discovered by Jean-Claude Risset, an expert in Physics and Music, while he was studying what makes instruments sound the way they do. It is suggested that the sound of a trumpet can be broken down into partial harmonics, which over a certain amount of time (attack duration of a time) evolve into a trumpet sound. Some harmonics are faster to reach the ear than others, giving the sound of a trumpet an exquisite and unique feature (see Figure 6). Important here is to note that multiple harmonics which behave differently contribute to how a trumpet tone is heard and perceived as a whole, which parallels the gestalt theory of the whole being greater than the sum of its parts (Metz-Göckel, 2016). This is to say that the separate harmonics that make up that tone are hardly separable by the human ear. Likewise, a single segment of a musical piece might not convey the same emotion as the musical piece itself. The ability to control spectral harmonics facilitates the process of simulating such sounds. In theory, it is possible to reach quite a similar trumpet tone if the right frequency modulation is set, as well as if the same amplitude envelope modulates the index of FM (Chowning, 2000). In controlling the single elements that make up a whole, it is possible to simulate the whole, and mold it into something else.

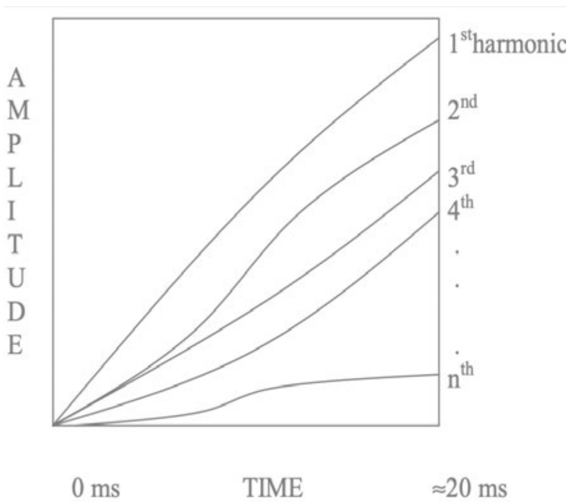

Figure 6. Harmonic Spectrum Evolution of Trumpet (Bader, R. 2013).

Language

System of Words for Spatiality

Spatial Language is a linguistic communication system that contains spatial cues and delivers spatial information or coordinates between humans. This can take the form of individual words or range up to sentences and even complex ideas. In language, it is important to differentiate between alternating perspectives, such as right and left differing for two persons facing each other (in this case left and right being opposites). The ability to process spatial language is connected to the ability to mobilize references and spatial perspectives, which is defined as cognitive flexibility. Single spatial words like “right” or “above” are egocentric pronouns while single allocentric terms usually encompass reference points, like north or west (cardinal directions). The use of spatial language can involve one or more spatial perspectives. Not only do language location cues exist, but the use of pronouns such as first, second or third person pronouns also specify perspective and location in reference to an object or event. Normally, in order to communicate spatial information, there is a "located object" and a "reference object". For example, in the sentence "the towel is in the bathroom", it is indicated that the towel is the located object and the bathroom is the reference location. This example is nothing but a combination of single spatial words constituting a more complex spatial description in the form of a sentence. Reference frames and perspectives play an important role in spatial cognition. It is agreed upon that reference frames are also to be found in non-lingual species, like newborns or rats, and that they are influenced by language. Reference frames are divided into *absolute frames*, *relative frames*, and *intrinsic frames*. The absolute frame refers to when an allocentric reference to a peripheral object or a surrounding is made, and usually uses the cardinal directions. For example, a tree or a street. A relative frame denotes the first person point of view, which encompasses the standpoint of the speaker. An example of that is: The book is left of the personal computer. This indeed would change when I, the speaker, change my position relative to them, therefore incorporating both references, mine and the objects themselves. Projective words make use of themselves in such a case, including left, right, front and so on.

This reference is egocentric. Lastly, an intrinsic reference frame discloses frames of things other than oneself. For instance, the garden is behind the house. It is object-based or other-perspective based, and also uses projective words. Researchers have different opinions on the wordage, nevertheless it is agreed upon that it is necessary to divide the reference frames into the three above-mentioned references. The use of the references however, interestingly differs in different languages and cultures (Waller & Nadel, 2013, p. 230).

A single sentence might contain all three reference frames. Fascinatingly, and despite vagueness in the interchangeably used terminology (left and right used both for relative and intrinsic reference frames) which in some cases might lead to communication errors, the sought after meaning is mostly understood correctly. Here is an example: *Look at the stars from the balcony over the street.* Looking at the stars takes on the *Earth-based* perspective which is allocentric and an absolute reference frame. *From the balcony above the street* is egocentric, indicating relative and intrinsic reference frames. This confirms the presence of cognitive flexibility in lingual species. Research has also yielded interesting results pertaining to when people are asked to give directions to their residential address. The subjects had used the word *you* (you go left, then right) when appealing to the listener, as they envisioned a conceptualized wandering, with so-called *Route Descriptions* that are language established, as reported by Linde & Labov (1975). These *Route Descriptions* (left or right) were, in the moment of explanation, based on the fabric of the envisioned perspective, and not the physical one according to Tyler & Tversky (1992). It is agreed upon that this *Route Description* linearizes space to fit the language's consecutive structure, since it is an imagined perspective. As noted before, pronouns also indicate which perspective is being portrayed. Consequently, using "the book is to the left" does not specify which perspective is being viewed, only the coordinates of the object in reference to a definite side. "The book is to *your* left" serves then as a perspective indicator, a clue that it is from your perspective to your left. These help minimize erroneous happenings in spatial language. In toddlers, during language learning and at around the age of two years, children can comprehend, use, and separate between *I, me* and *you.* This is an indication of differential perspectival understanding, as the word *me,* based on the setting and who the speaker is, can be used for different perspectives (Waller & Nadel, 2013, p. 233).

In the physical world, it is rather rare that our environments are simple and only produce a reference object and a located object. It might be so, that the mentioned "*the book* *to your left"* example, also has other reference points for the listener. If the book is also beside a flower vase, a cup of coffee, and car keys, there is a process of evaluation that happens before the chosen reference frame is decided and assumed. The process can happen in both serial and parallel processing, meaning that the processing of multiple reference frames happens simultaneously or consecutively. Simultaneous processing is linked to higher cognitive flexibility. Both processing applications have advantages and disadvantages. Duration is also variable, as serial processing takes up more time than parallel. Research also strongly backs reference frame mobilization Waller & Nadel (2013).

Waller & Nadel (2013) report about a research by Hauward & Tarr (1995) where spatial language processing and the processing of spatial location in perception employ the same structure in the brain. This bridging of processing between spatial language and in the case of this research, vision, proves that spatiality can be crossmodal. Waller & Nadel (2013) write:

> Brain imaging data have suggested that spatial preposition processing in language and spatial location processing in perception engage in the same system (Hyward & Tarr, 1995). This research has shown reliable processing of projective prepositions in the left supramarginal gyrus (SMG) regardless of whether the comparison modality for the preposition was language or pictures (Damasio et al., 2001: Noordzij, Neggers, Ramsay, & Postma, 2008). (Waller & Nadel, 2013, p. 233)

Circumstantial elements certainly influence reference frames. The interpretation and assumption of a reference frame depends on what is being located and referenced. A sphere, in contrast to the shape of a human or a cube, does not have sides. Therefore, it might be confusing and perspective-dependent if left or right spatial words are used to describe the location or reference. For example, "*the player is behind the sphere"*. The sphere itself lacks sides and does not have a front or back side. A human, although arguably having a multitude of sides, has axes which indicate front, back and sides. Therefore, contextually, describing a located object as a reference point to the human body automatically assumes the perspective. Note that this is different with above or below, as these are bridges over the three frames (absolute, relative, and intrinsic), but assume almost definitely the absolute frame in regards to gravity. There is evidence as well that when describing a location or a shape, having common ground with the listener, facilitates the delivery of further information from and to the listener's perspective. Having the same language between the communicator and the receiver also facilitates the transmission of spatial concepts. This refers to, for example, using references that are known to both parties that aid in the acquisition of new information (Waller & Nadel, 2013, p. 234).

Spatial language translates to a physical activity influencing perceptual and cognitive processes, hence guiding a part of life. For instance, given directions to one's home become instructions for the receiver, with which the route is imagined, assumed, and executed. In turn, not only would spatial cognition be important for imagining, analyzing and recalling directions, it would also be important for analyzing the where and how of the current moment spatially, in order to apply the instructions and succeed in employing these spatial descriptions. Two theories have been proposed for the mental reproduction of space through spatial description. Waller & Nadel (2013) suggest the following:

> The first is the view that spatial memory is composed of amodal, symbolic representations, such as a cast connectionist network that codes for and reflects spatial relationships. The second, more recent view adds a substantial modal, analog, and sometimes image-like component to earlier theories, and language-processing researchers have begun to apply this approach. (p. 238)

This suggests that spatial cognition's meanings are symbolic, and can be imagined images (Waller & Nadel, 2013, p. 238).

Sound, Space, and Time

Sound in Space and Space in Sound, over Time

What is space? According to different philosophers, engineers, psychologists, artists, mathematicians, physiologists, and physicists, it can mean a multitude of different things. In spatial cognition and with regards to internal reflections of space, personal space has been defined as the space within arm's reach, while public spaces are somewhat more common and enable entities to move within them. In music, space and spacing may indicate the distance between one note and the other, whether concerning frequency or duration. Space can be represented as mental capacity or any other hosting capability. For installation artists, space may indicate the distance between sculptures and the characteristics of the environment they find themselves in. It can reveal emptiness or fullness and other forms of symbolism. Furthermore, space can also be substantiated according to a set of rules, like boundaries within it that form acoustics, where performers and audience have venues defined by such terms. The evolution of concert halls is such an example. The arrangement of boundaries, electroacoustic elements, stage, performers, the audience, and the bodies that move within it constitute a space. For sociologists, space is defined by its purpose as well as by the inhabitants of the space and how they move within it and use its objects e.g. municipal and residential buildings, roads between them to make movement possible, both of which are spaces that have different properties adapted to their usability. Place can be attributed and so constructed, that it serves a purpose, and only then can it be considered a space, according to Martina Löw. She also defines the terms *spacing* and *synthesis*. In a sound installation, a place does not become space, before *spacing* (the arrangement of its elements), and the presentation of the art within it occurs, prompting *synthesis,* which gives a place its value or meaning, namely making it a space (K. Sharma, 2016).

There is a difference between space and its representations. On the one hand, space consists of the three axes: Width, Height, and Depth. On the other hand, its representations can take on different forms,whether it be placement of objects in space, or the interaction between them. An intuitive example of reflections in the world of acoustics is reverberation, which is the "illumination" of a space in which a sound takes place. It is a result of the cooperation of the exciter and the excited , and not of the space itself. To make an analogy, the light that shines from a candle lights up a certain place (room) as a result of the reflections on its surfaces (and the properties of the surface, like what wavelengths or color they reflect), and not of the candle or the space alone. Therefore there is a difference in the representation of space between moving sound over a certain distance, that being a physical, spatial property of size, and between the placement of a sound in a space that is illuminated. Another example of a representation of space is delay or echo, which is a form of relatively late reflection of sound on a surface.

Sound can only exist in space and time, as it travels in space through matter and in time. Music specifically has a cultivated link with space and time. Early humans chose certain chambers for certain activities, while disregarding others, due to particular resonances. This binding is neither causational nor serial, but rather intertwined and mutually influential. A long reverberation decay time sustains a note which then perhaps overlaps with the new note, forming a new sound or chord, an effect that might or might not be desirable to the corresponding musical context. The words of a preacher in a church come across as much more striking when the early reflections and echoing decays emphasizes them. This is a clear example on how space dominates and imposes a condition on the composition. Interestingly, it is evident through history that the creation of space and sound was never simultaneous. It is rather a result of trial and error. The experimentation with space and sound throughout history has generated information (whether physical or psychological) that in turn influenced the production of sound and space. It is a cyclical relationship of doing, evaluating, and reforming (Blesser & Salter, 2009, p. 128). Blesser & Salter (2009) stated that pop music subcultures are often exposed to different spaces in the same musical piece which permits the unorthodox use and perception of modeled and synthesized spaces. Space representations don't have to sound consistent, nor do they need to be accurate in their reproduction of space. The multiple space usage implies that the listeners also find themselves in multiple spaces while actually not changing places. Spatial consistency becomes itself an artistic attribute and a creative decision that the aural architect determines. Jazz and classical music lovers tend to exclusively listen to music in concert halls, as the space plays an important role in how the end product sounds in this specific space.

„Sounds signify events taking place: Babies crying, brakes screeching, birds singing, people talking, and water falling. All sounds are the result of dynamic action, periodic activity, vibrations, sudden impacts, or oscillatory resonances.“ (Blesser & Salter, 2009, p. 15). The reveal of sonic information aids the brain in decoding a singular happening, hence its acknowledgment and evaluation. This has played a role in the survival of human beings. A sound heard is a revelation of the existence of the object that caused it. This is perhaps the reason why sounds provoke the turning of the head. In many cases, the perception of a sound-making object can trigger a certain response. Startling someone tends to trigger the fight-or-flight response, which makes for an involuntary reaction important to survival. With time, associations occur between sounds of objects and meanings, and like Pavlov proved

(Lattal, 2013), in a conditional way. Pavlov was able to trigger an innate reaction of salivating in dogs, with a sound that did not resemble that of food. The formation of new associations is therefore possible with simultaneity of events and utilization of older associations. Pavlov proved that with associativeness, the transference of innate behaviors is possible onto something that is new or rather not innate, as if it were of the same importance. This is an example of classical conditioning in which Pavlov proved that dogs, much like humans, learn associations over time. Brech & Paland (2015) mention Peter Doyle, who researched the use of reverberation and delay or echo effects in music in the 1960s, and emphasized that these effects are broad and do not fit an individual connotation, but are rather "rich associations with space and place" (Brech & Paland, 2015, p. 148).

Using aural architecture as a maneuvering technique is a capability developed by humans. As a given, surroundings are experienced through senses. Each sense occupies a different dimension of the same object. For a sound to be heard, there must first exist a force (energy) or an action, as well as an interaction of matter, which takes place through time. The use of echolocation has been proven to be helpful for people without sight (Blesser & Salter, 2009), which is based on the interaction of a person's own body, as well as other sounds sources, with the environment to gain information about it. Imagine the difference in acoustics by standing 10 cm away from a wall in contrast to 1 meter away (distance acknowledgment using reflections). Similarly, the pulling up of a train might influence behavior and provoke the taking of two steps backwards.

The emergence of electroacoustics in recent history made decoupling sound and space possible, which yielded new aural perspectives. Listening to music recorded in a concert hall with a characteristic reverberation time of five seconds is possible in a living room of different reverberation characteristics. While the reverberation doesn't match the listening environment, the effects of the recorded reverberation still apply to some extent, not regarding the combination of the recorded reverberation and that of the living room. Today's headphones provide yet another dimension of aural perspective. Binaural sound displaces the listener into a new environment. The aforementioned physical restriction tends to vanish in the realm of spatial synthesizers, leaving this restraint of physical origins solely to the sound designer (Blesser & Salter, 2009, p. 130). Blesser & Salter (2009) indicate:

> Such technology presents aural architects with a clear choice: to duplicate a specific aural experience modeled after a real space or to create an artistically meaningful one in a virtual space without a physical counterpart. This choice has important artistic implications. (p. 130)

Here, the distinction is made between the physicality of space and its perceptual experience. Spatial replicators or synthesizers with manipulatable outcomes offer an illusory experience, and a physical language of something virtual or of unnatural origin. Replication of space simulates an external attribute of an experience and calls up its internalized effect (Blesser & Salter, 2009, p. 131). The apparent emotional effects of hearing the stimulus are individual and possess, to an extent, independence from the physical stimulus. It is an internalization of an external surrounding or object. With the use of virtual spatial synthesizers, the physical experience becomes symbolic and is transformed into an algorithm (digital) evoking different and even new impressions arise. The appeal of these effects, characteristics, and order are aspects of choice of the sound designer or in this case, aural architect. Nowadays, switching between modeled realistic spaces and virtual spaces in spatial synthesizers can happen seamlessly, as the two extremes lie in a continuum.

It is helpful to make the distinction between the two classes of spatiality. There is, on the one hand, sound in space, which is when a sound as an object moves within space. For instance, delaying a sound over a four-loudspeaker setup with different delay times produces a certain trajectory, or simply forms a trajectory with amplitude panning. Space in sound, in contrast, is rather when a space is “illuminated” and made hearable. This is, for instance in a four-channel setup, when a sound excites global, four-channel reverberation (single- or mixed-channel illuminations). The main difference here, is that the first is the use of space as a trajectory utility, manifested according to the loudspeaker setup and the dramaturgical movement, depth, and velocity, while the latter is concerned with the creation of space that possesses fixed or alternating qualities (direct to reflected signal ratio, boundary material, size and shape) (K. Sharma, 2016, p. 69).

Reverberation is one of the most pronounced spatial attributes regarding sound in enclosed spaces. Blesser & Salter (2009) mentions that space itself is an extension of the instrument once it sounds from within, just as a resonant body of a violin is an extension of the vibrating strings. Not only does it amplify and spread the sound, it also adds coloration to the frequency spectrum. The same concept applies to a space, e.g a concert hall, where the performance space fuses itself with the instrument, creating a meta-instrument. This also leads to, through reflection and dispersion, certain resonances being amplified and/or attenuated. Although reverberation is made of various reflections, they fuse into a single constitution to create a single body. In a four-channel format, the aural architect has control over what perspective the audience takes. As it is not possible to fit the public in a regular sized violin, they must be seated in the second resonant body, the concert hall. Using spatial synthesizers though, one can separate the resonant body from the exciter or even have multiple resonant bodies.

In composing spaces, sound designers may plan, construct, erect, break down, bridge, and move elements in the sonic space, much like urban engineers do. Space is created for the function that it serves. Overlapping multiple spatial layers in a four-channel installation setup makes a constellation of different spaces with different functions. For example, a sound which is bathed with a global reverberation moving in space indicates that the movement takes place within an enclosed space. A second layer is added to that, but it is within a space of different diffusion characteristics. That communicates a second and different space. A third layer of sound might have no reflections, and only appear from one constant corner, giving significance to occupied and non-occupied space. A fourth layer might again be added, revealing other characteristics like acoustical absorption or outdoor quality.

In an enclosed space, such as a concert hall, where conventional concerts are featured, two firmly connected concepts take place. The first is as follows: if any sound source in a concert hall is taken, for example percussion, it produces direct sound waves in a circular fashion (Blauert, 1997). These sound vibrations propagate in a "random" or "all-direction" manner, until they hit a surface, barrier or object. Although the sound source ceases to create sound after the impact, the sound still extends after it. This elongation or extention of the sounding object in time is called *temporal spreading.* A second type of spreading is called *spatial spreading*, where waves of a sounding object arrive from not only the position of the sound source, but from a greater area (Blesser & Salter, 2009). Reflected waves that bounce back on different surfaces arrive at the ear of the perceiver, indicating different directions of propagation. The word diffused sound is used to describe a reverberation quality, as reverberation is an iteration of the location itself due to an exciter. Note that temporal and spatial spreading happen simultaneously, but are rather different concepts that can be handled separately as design elements. This becomes an artistic means of evoking an effect, in the exploration of sound in space using panning techniques, and space in sound using simulators. Envelopment is an attribute which is defined by time and space as well, and is enhanced with spatial and temporal spreading, as well as spatiality in general. An elongated and diffused note has a greater envelopment than a sound that is short and resides at one point in space (Blesser & Salter, 2009, p. 134). Thomas Lund indicated in a lecture at the Tonmeistertagung 31 in Düsseldorf the difference between immersion and envelopment. Immersion works in an outdoor situation as well, when hearing a surrounding that is occupied with sounds from different locations, distance, and direction (forest or a flat field). It tends to be characterized by direct sound in contrast to envelopment, which makes use of reflections to *enfold* the listener, and is usually characterized by indirect sound. This is not aimed, though, at strictly defining direct sound as unenveloping or indirect sound as unimmersive. Usually, it is also the case that a combination of both is used. Lund also included that to sense something it must be active, experienced and undergoes change. He reports the importance of decorrelation even in frequencies as low as 120 Hz, and of reverberation, as his decorated recorded signals yielded more immersiveness and envelopment in contrast to correlated signals.

In modular synthesizers, it is possible to combine spaces of different spatial synthesizers creating a combination of spaces in a single hearing space. Crossmodally, this makes little sense, as listeners find themselves in a single place, but indeed, multiple spaces are heard. An intriguing concept which happens using modular synthesis with space is the composition of acousmatic images in space, and then the manipulation of these images to produce a sequence of meaningful events. Like film editing, the sequence of events can change the meaning of the information being communicated. So the images composed using space can unfold into acousmatic events that are meaningful. For instance, having one sound, such as an ambient pad, move from initially a single loudspeaker and spread to all four, tells a story of envelopment, growth, and expansion. In contrast, when sound is generated from all loudspeakers and moves to one of the four, this implies shrinkage, reduction, or compression. These metaphors can be arranged to build images. The transformation of the images into others makes it possible to create a story, a whole structure of information that would be rather lacking in information without its spatial context. These metaphors and images might or might not have reference to real life situations or be traced back to experience, but can still establish a ground to build on. Like the conditioning of Pavlov, one can use pre-existing metaphors to signal newer ones. Trever Wishart (Emmerson, 1986) mentions that the expression of the affiliations of events evolves into a symbolic structure. In defining a sonic landscape, he argues that the landscape of a concert hall is an orchestra sound (playing instruments, seating movements, audience). Yet, this also applies to orchestra sounds coming out of loudspeakers. The objects of action in these cases that produce the landscape are different, whilst the landscape stays the same. Hearing a sound and not being able to locate a conceivable source, induces a sense of disorientation, which renders it useful, for instance, in prompting a sense of strangeness and alienation in science fiction environments. Giving the unknown a contextual sense makes for significant meaning. Furthermore, a landscape can be divided into 3 independent parts, namely the "nature of the perceived acoustic space", the "disposition of sound objects", and finally the "recognition of individual sound objects". To briefly explain, the first refers to the establishment of a scenery, an arena, in which sound-producing events take place. This includes all kinds of spatial cues, like spectral cues, reverberation, echo... etc, which reveal the characteristics of space. The nature of the landscape can be determined by a set of rules, one of which, for example, might be reverberation time. In that case, single sonic objects react differently to these predetermined sets of laws as their sonic qualities change (for example, pitch: a low pitched drone might have less spatial cues than a high pitched percussive sound). This becomes a product of the illumination effect of a space by sonic objects. Pierre Boulez's idea of space in sound and music shows: “observed by the surrounding sound, or we take a listen to the sound from a distance, observing it from the outside” (Harley, 1994, p. 160). Boulez means to clarify the perspective change between the sound and the listener. For instance, a quiet sound with rolled off high frequencies might indicate that the listener (or the object) is at a great distance, partially defining space. For humans to recognise the meaning of an unknown (and sometimes even known) sound and its qualities, it usually has to be heard in a contextual manner. A bare-faced ibis for example might not be easily identified until it is heard within the context of a tropical landscape. Reproducing cricket sounds electronically might deceive a listener, when heard independently, into thinking it is a real cricket, but is instantly recognised as synthetic when contrasted with and compared to other cricket sounds. It is also useful to mention that individual recognition of sound objects partially relies on crossmodal correspondence. The transformation of sound into another kind of sound is somewhat more ambiguous than visual elements. The transformation of a pulsating white noise into waves lapping on a beach, for example, tends to have a whimsical effect. This is of course different when discussing abstract or realistic sounds (or a combination of both), as in the mentioned example, moving from white noise which is a rather unnatural sound, into a sound of similar spectral features (sound waves). This transformation in spectral content over time is called the morphology of sound, as described by Denis Smalley. A sequence is not only present when creating images that achieve a metaphor, but also in the single sounds themselves. Figure 6 shows a somewhat organized sequence of trumpet harmonics over time. Furthermore, rapid and random change of periodicity in sound fills the spectrum and results in noise. On the other hand, a sine tone is rather the opposite, in that it occupies a single band of frequency that has a steady periodicity. Modulating qualities of sounds can morph them to other, new sounds. Groupe de Recherches Musicales (Emmerson, 1986), categorizes continuity of sound into three types. The first is discrete, a category that is typically transient-like, and with very little modulated qualities, which can be resembled by an impulse. The second is iterative, the constant sound of multiple impulses rapidly struck. The third is continuous, which as the name suggests, is a sustained tone, for instance, a tone provided by an oscillator. Furthermore, there is a difference between a synthesizer sustained note, which generates its sound from constant electrical supply, and the sustained part of the sound of a gong instrument. The first is continuous because of its electrical supply, is steady and has a regular periodicity with a constant spectromorphology, if it is not modulated, while the other is rather the opposite, because of the resonant property of the gong´s corpus, and the inevitable decay of the amplitude after the impulse. This difference in continuation is respectively imposed and intrinsic continuation. This segregation helps in the identification of the nature of the sound as well as its gesture. This is to say, for imposed continuations, there is a need for "excitation", where synthesizers do not need intervention or excitation to proceed . Interestingly, continuous sound, like the one that comes out of an oscillator, did not exist before the development of technology. Refrigerator sounds are also a good example of that in daily life. Another difference between intrinsic and imposed continuation is the ability to perceive modulations and variations of the sound with intrinsic continuation, while with imposed continuations, modulations and fluctuations are less perceivable. Here, and importantly in modular synthesizers, one has the ability to have control over the fluctuations and modulations, the amplitude envelope, and on top of that, the sound's spatial position.

Concerning Denis Smalley's spectro-morphology, it is the view on sound in which the frequency spectrum and its evolution through time is taken into consideration. It is a proposal in which almost all sounds can be observed and analyzed. The spectro-morphology of the human voice, for example, is extraordinary, because it is able to produce a multitude of different pitches and spectra, like vowels and consonants. The gestural nature of a human voice is translated into sounds, including instrumental music. Recently, electronic and technological advancements, as well as the power of signal processing using digital technology opened up a wide variety of opportunities for the detailed modification and manipulation of sounds, lending human gestural qualities to rather subtle and dull sounds (like imposed oscillator sounds). Denis Smalley writes: "As sound sources, however, language and vocal sounds have discovered a new significance as a result of their contact with electroacoustic media" (Emmerson, 1986, p. 62). So, the animation of rather inanimate sound objects can be preserved with gestural modulation. In the analysis of sound, so-called reduced listening plays a significant role. Reduced listening is the decoupling of a sound from its source, and looking purely at its independent properties objectively. This helps in disconnecting the source from its meaning or cause, making it rather more objective. This is analogous to saying a word consciously and repeatedly, until it loses its meaning and is reverted to an unconditioned sound. In such an approach, we can view all sounds as concrete or abstract, for sounds can hold both qualities. The concrete qualities are experienced in life, whereas abstract (or semantic) qualities are indefinitely learned. These categorizations are dependent not only on the composer, but also on the receiving end, the listener. Sounds engaged in mimesis almost always carry an abstract definition. Here again, broader abstractions can be made with the sounds's structure, as well as its spectrum and the combination of multiple sounds. It is worthy to mention that these semantic qualities or functions can be newly formulated, since they are learned.

In explaining a piece he composed called Kontakte, Stockhausen (1972) refers to the tone which is sustained for almost four minutes as the tone that is constantly there in the spatial sense. Other tones and sound effects only burst in rapidly for a matter of seconds and then fade away. The sustained sound in this case serves as a reference "anchor" for the perception, as other sounds of different durations and spatial location relatively "pass by". A reference sound was first established, and then the appearing and disappearing of other tones add to or subtract from the concept of what is occurring and how it is changing. Stockhausen follows the pitch of the sound hand gestures, indicating *spatial height* and *movement*. Furthermore, in Kontakte, what seemed to be a tone drops in pitch and becomes the sum of periodic impulses. Additionally, changes in intensity are referred to as "going away very far" and then "coming again". Stockhausen caters to the relationship of time and space as well. There seems to exist a middle position of orientation in reference to speed, which is the equation of space and time. Humans, for instance, are bound due to the physical nature of the body, as to how fast or slow it is possible for a movement to be executed. A middle ground for speed of orientation exists, which is between fast and slow, having both in reachable distance (Stockhausen, 1972). This is also evident in music, as music tempo and pitch are adapted to the speed of perception. There is certainly a point when music becomes too fast, and merges into a single sound, or too slow, where the whole is not perceivable anymore as much as its single parts. This is a result of technological advancements that changed the human perception of time. Speed can now take on different perspectives and therefore the relative perspective has been assumed. Whatismore, take an example where a car passing by a person on the street, seems to race by very fast in contrast to looking at a plane from the ground. From an absolutist perspective, the plane moves at a much higher speed than the car. From a relativist perspective, the car in this case, due to its closeness to the subject, seems to move at a higher speed than the plane. Relative to each position, different speeds can be assumed, although from an absolutist stand point, both have a definite speed. Playing a sound in reverse is another example of the change of time direction. Similarly, looking down from an airplane, everything seems to move incredibly slowly. This effect of technological advancement changes our middle position of orientation.

Iannis Xenakis, who is known to be a space componist, differentiated between time and duration. While time is the ongoing and constant flow, duration is a limited window of units used to count time, which can be moved around in time. It is the span between two degrees within time. Both a short staccato and a long sustained note exist within time, only their durations differ. Time is an abstract phenomenon that we as humans do not possess an organ to sense like, for instance, light. As time is a natural phenomenon and a condition applied to everything, it is unstructured by nature. Humans are indeed attracted to the idea of structure, as unexpectedness might induce anxiety (like fright because of a sudden, loud sound). Music, for instance, can be thought of as structured events in time that play on expectancy.

Denis Smalley (Emmerson, 1986, p. 65) points out that sound can not be experienced outside of time, whereas time is experienced through spectral motion. Change, therefore, can only happen with time, where comparison between before and after can take place. Furthermore, the facilitation of space according to spectromorphology can be put into five categories: *spectral space, space as time, resonance, spatial articulation*, and *the transference of spatial articulation. Spectral space*, which is to be described as a technique later, is the design of space using the spectrum of sound, where the harmonic and inharmonic spectrum are laid onto the space, dividing and categorizing it. This can also mean that space is encoded into the waveform, e.g. the manipulation of a signal in order to make it appear to be positioned somewhere in space, like the psychoacoustic effect in pitch-height or the elevation effect. In *space as time*, changes in spectro-morphological design, like the harmonic structure of a sound, can only be unfolded and perceived through time. Without spatial movement, sound can still present space and time within it, like the above-mentioned pitch-height effect, as well as the distance between notes or intensity of sounds (how close or far). This does not require immediate space, but does require time to move through, hence space as time. *Resonance* is widely common, which is the effect applied on sound in enclosures. This effect deforms a sound's structure, rather than creating something new. Resonators, springs, and plate reverberators are examples of an extension to the original sound of potentially different characters, which affect the *internal* structure of a sound. Spatial articulation is the *exterior* element of sound, like reverberation. Spatial articulation is defined by properties of space, e.g. what dimensions it possesses. Furthermore, space can be realistic or illusory. Sound Gesture describes trajectories throughout space, while texture of sound can be distributed onto space (like the movement as causation to spectral change). Having the same sound go on a trajectory through multiple spaces in which sound reflections occur, would be perceived as the change of space for the listener (Emmerson, 1986, p. 91). Lastly, the transference of spatial articulation, is where the end product is heard, whether in a living room, through headphones, or in a professional studio environment. This once again affects the perceived spaces and morphologies of the sounds of the piece being listened to.

Elizabeth L. Anderson mentions (Brech & Paland, 2015) that there is a reference to how spatial sound is perceived. In an attempt to analyze *Sirius*, a work by Stockhausen, she writes that there exists a "Reception Behavior Framework", which she based on François Delalande's listening strategies. Here, perception of spatial sound is categorized in four different classes, none of which is superior to the other, but which are rather different in their nature. The classes are the following: *sonic properties, structural attributes, self-orientation, and imaginary realms*. These classes go beyond the definition of their names and dive deep into abstraction. *Sonic properties* is the type of listening in which the focus is lended to the sound source and a description or analysis of it arises. *Structural attributes* take into account the universal form of the piece. This can also include abstract concepts existing in the work that shed light on its structural design and arrangement. *Self-orientation* is the reception that lends the background and psyche to the work, in which reflection and emotional engagement is present. Lastly, *Imaginary realms* is the reception that journeys outside of the realistic world and comes closer to science fiction and fantasy. Although these categories are seperated, a listener can indeed experience multiple reception behavior, called *hybrid* listening or skip from one reception to another , called *dynamic* listening. A *combination* of these listening behaviors is rare but existent, where a listening goes through hybrid, independent and dynamic patterns. These listening behaviors are nevertheless completely individual and person dependent. As to the reception of any impulse, its symbolic meaning resides in the receiver (Brech & Paland, 2015, pp. 259–263).

Movement, Directionality, and scale

Humans tend to identify with forward being the main direction of maneuvering. This directionality is clear to identify in Greek amphitheaters, where everyone in the audience looks in one direction, at the place where the theater piece takes place. Today, concert halls and theaters still have the forward direction.

In audio, traditional forms of recording techniques have preserved the soundstage of events. Recently though, this has not been the case in somesound art, where the concept forward directionality is loose and all directions are possible (Brech & Paland, 2015, pp. 15–18). Brech & Paland (2015) writes: "Recent installation and other open-work formats challenge this fundamental exclusion: an audience member may be free, after all, to ›wander into‹ the work!". Furthermore, the scale in which sound is heard in real life using technology has changed. Brech & Paland (2015) introduces a diagram that helps in facilitating space from a sound perspective (See Figure 7 and 8).

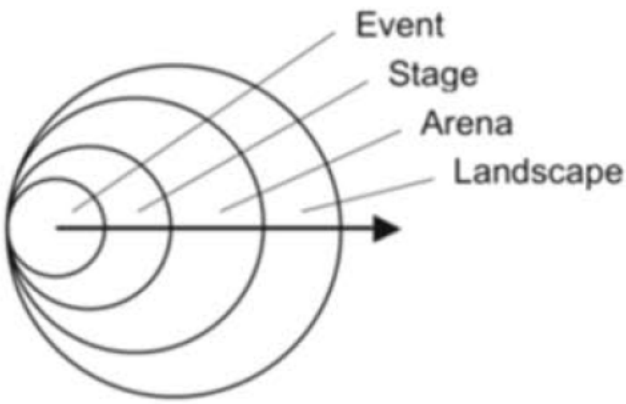

Figure 7. Brech, M., & Paland, R. (2015). Directionality "Forward"

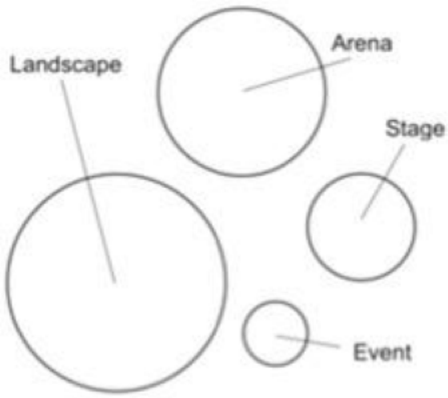

Figure 8. Brech, M., & Paland, R. (2015). Fragmentation - Dislocation

The Event is the physical source of the sound. The Stage is the spatial organization of the sound and its spatial reference to other events. The Arena is the environment and the response of the environment to the sound. The Landscape is the horizon which the sound can reach. Beyond that, it is now possible to amplify sounds that are barely hearable, or attenuate other loud sounds enough to be listened to repetitively. Sound intensity is not bound to the energy generated by the event itself anymore, but can be handled separately, changing the nature of the sound, and effectively, the psychoacoustic representation of the sound source (see Figure 9). These boundaries between what Brech & Paland called "Event", "Stage", "Arena", "Stage", and "Landscape" dissolve, rendering the alteration of their representative characteristics flexible. This in turn frees sound designers to find new ways to experiment with space and its fragments. This establishes the power of technology in magnifying and de-magnifying the dimensions of space and time. By eliminating the sequential propagation of sound, so too is its directionality eliminated. Hearing the landscape before hearing the event reverses the natural propagation direction e.g. listening to reverberation and then the sound source. Thus, we can assume that the field of sound propagation is a round sphere, where all sides are of equal distances to the centerpoint. Brech & Paland (2015) also point out that because of the new technologies of telecommunications, recording devices, and even the advancements in astrology, new limits have been added to the four earlier boundaries, namely those of "telecommunications", "history", "memory", and "extraterrestrial" (Brech & Paland, 2015, p. 24). This explains the concept of up- and down-scaling space and time using telecommunications, the emergence of non-linear time, and the imagined beyond-earth sounds of extraterrestrial objects.

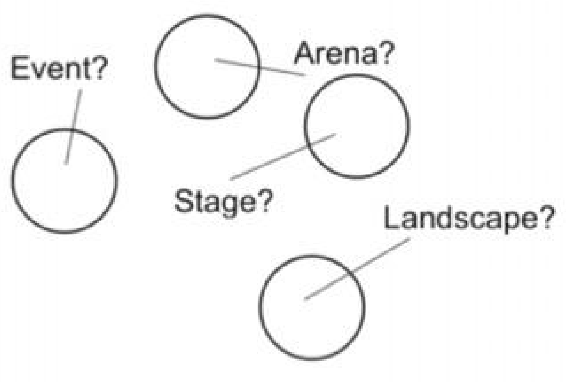

Figure 9. Brech, M., & Paland, R. (2015). Space Frames - Change of Scale

Motion is evident not only in sound movement, but also in single or multiple fluctuations of different aspects of sound. Therefore, one can say that movement is not only in the physicality of space, but also in the spectrum of sound objects. It has been established that, in nature, very few or no sounds at all have a steady state. Differences in amplitudes of frequency bands are present in even the slightest and shortest burst of sound. It is therefore of importance to take a look at these sounds and their relationship with spatial motion. Pitch, according to C. C. Pratt (1930), has a spatial character, which is defined by the term *Pitch-Height,* where high pitched sounds have a higher position than lower ones. Stockhausen (1972) uses hand gestures to indicate the pitch changes of sounds in his piece "Kontakte ", where higher tones were gestured in upward hand movements and lower ones with downward ones. Even the terminology of using high and low as sonic descriptions originates from the spatial domain, due to the lack of terms for sound. Thus, spatial attributes are already innately assigned to distinct parameters and characteristics of sounds. Having control over these single characteristics and their quality lends creators the possibility of achieving spatiality that would be mirrored or represented in physical spatiality. It was not so long ago, in the twentieth century and with the rise of multichannel audio setups, that sounds were granted the ability to move in literal space. In The Language of Electroacoustic Music (Emmerson, 1986) (see Fig. 10.), Smalley makes a distinction between the essential motions: *Unidirectional, bidirectional, reciprocal, centric/cyclic,* and *eccentric/multidimensional.* Although these are used to describe motions of spectro-morphology, they suffice to describe physical sound movement. This is relevant, as in modular synthesizers, control over spatial position and inner sound fluctuations can happen simultaneously, where sonic qualities and physical motion can interact (mirroring or opposing each other). e.g. pitch rises while sound position descends.

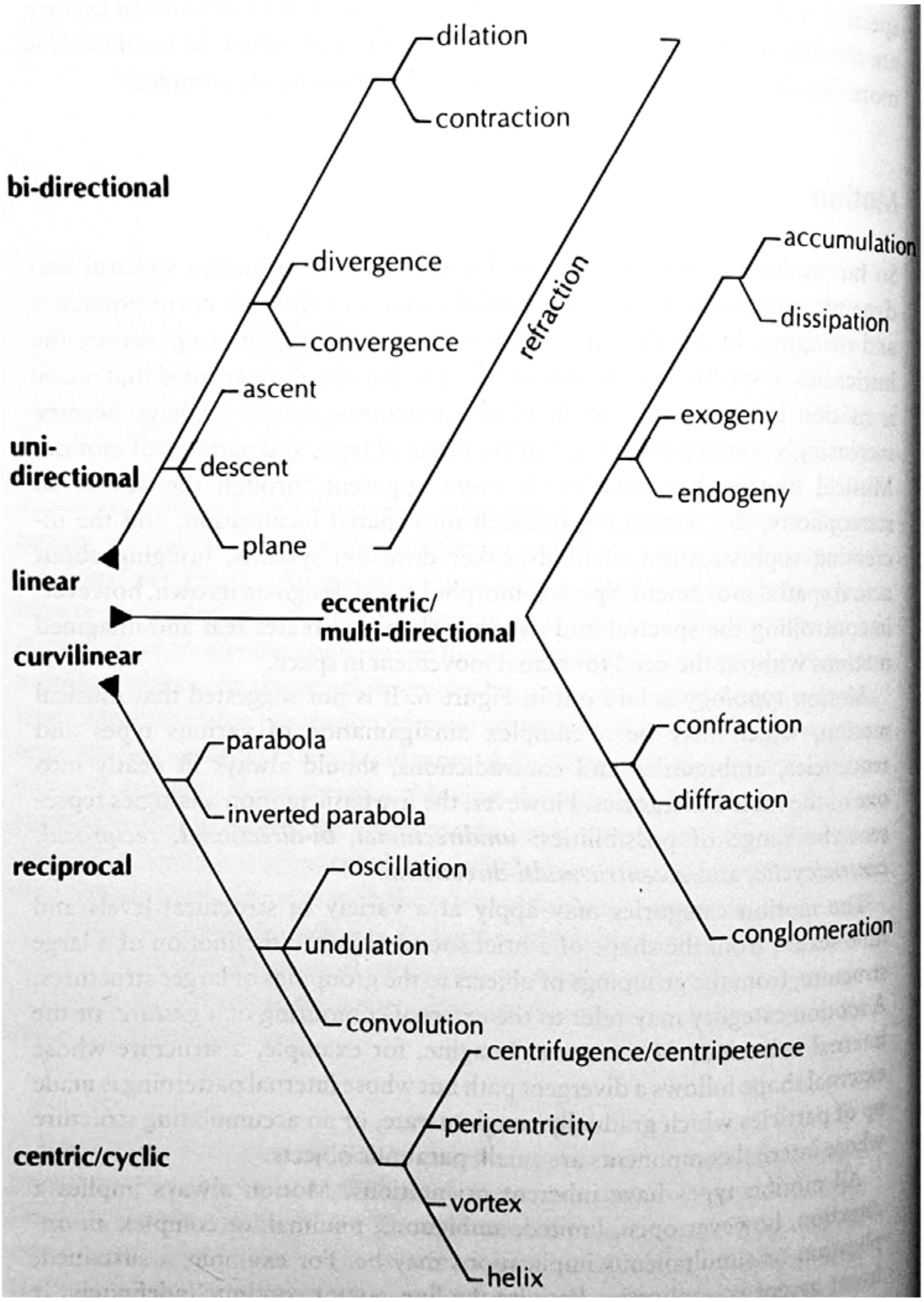

Figure 10. Emmerson, S. (1986). Motion Topology.

One thing is definite, all kinds of movements have an orientation, no matter how small or large, fast or gradual, directed or diffused they are. The following is to be understood under linear motion: *unidirectional* and *bidirectional*. *Unidirectional* motion, as the name implies, has a single direction. It involves motions like *ascent*, *descent*, or *plane* which are, respectively, rising, falling, or straight motions. *Bidirectional* on the other hand encompasses movements that occur in two directions. This includes *convergence*, which is the motion from the outside inwards, as well as *divergence*, which, in opposition to *convergence*, is motion going from the inside outwards (like the phenomenon of water ripples created by a water drop). Bidirectionality also includes *contraction* and *dilation*. *Contraction* is the subsequent reduction in size of an entity and *dilation*, in contrast, is the enlargement of an entity. *Refraction* is the change in direction and nature of motion from *linear* to *curvilinear*. A three-stage envelope is a *curvilinear* motion. A three-stage envelope such as the attack, sustain, decay envelope encompasses three motions. Motion is *reciprocal,* if it is countered by an opposite motion. *Parabola* (curve motion) and its inversion, *oscillation* (motion between two distinct locations), *undulation* (continuous movement similar to sea waves, which are not duplicates) and *convolution* (intricate intertwine of symmetrical motion) are all part of reciprocal *curvilinear* motions. *Centric* motions, also called *cyclic*, are motions that orbit a single point. *Centrifugence*, which is the opposite to *centripetence*, describes sound motions that move away from the center. *Pericentricity* describes motions around the center. *Vortical* movement is defined by a central axis, around which objects or sounds move. *Helical* motion indicates spiral movement, which in turn involves acceleration and decceleration. Lastly, *eccentric* or *multi-directional* motions are an extension to *convergence* and *divergence*, both being movements that can be morphed into *eccentric* motions. *Exogeny* is the expansion and attachment to the outside periphery. This can be imagined as a point of reference with magnetic tendencies of attracting objects to it, gaining in size. *Endogeny*, in opposition, describes the expansion of an entity from within (like the thickening of a tree trunk), so that it grows and becomes more dense. *Accumulation* and *dissipation* are respectively, the increase or decrease in motion over time. *Confraction* describes taking an entity apart into different increments, while *diffraction* is the breaking up of an entity into a composition or a structure. Last but not least, *conglomeration* encompasses the coming together of several entities to form a greater structure. These different types of motion can be singled out, broken down, overlapped, and used together to create even greater and more complex forms and motions. These differ from other types of movements, in which a point of reference or a "forward" is not established, and motion is free to happen and be interpreted differently, yet maintains its core laws. Several motions can also have definitive traits that separate them from other motions. A group of motions can be *synchronous*, *continual*, *joined* or *periodical*. *Synchronous* movements occur simultaneously and are, in terms of synthesizer language: clocked, while *asynchronous* sounds do not correspond to each other, moving freely. *Continual* movements, as the name indicates, are sustained and endured for a distinct period, while *discontinual* movements are movements that experience similar abruption. *Periodic* movements are repetitive and oscillatory, while *aperiodic* movements are ever changing, not having a period or a fundamental frequency. Moreover, groups of motions can be classified into further categories, like *locked* motion, which is a group of motions that concur and are interrelated. These sounds that make up the flock do not have to be synchronous, but have commonalities with each other, which makes it possible to identify attributes that mark them as related. Similar to flocked motion, *streamed* motion is a group of motions that are intertwined, but unlike flocked motion, the individual motions are still discernible. They have an established commonality as a group, but are easily identifiable as single elements. Finally, *contortion* is defined by the disorganization of movements that avert both flocking and streaming as they overlap, and can hardly be singled out.

The described movements can be translated into voltage and achieved in the modular synthesizer context as follows: a *unidirectional* motion, like *ascending* or *descending*, can be acquired through a low frequency oscillator ramp or decay voltage (or a triggered envelope). A *unidirectional-plane* motion is a held voltage offset. *Reciprocals* like a parabola can be achieved with a modulation source and a negative inversion of the positive output voltage. Wavefolding can also be a *reciprocal*, but unlike a parabola, it differs when the symmetry is changed. *Cyclical* motions, like *rotation*, can be represented through sine wave oscillations of a low frequency oscillator. Spin or spiral cyclical motions can also be achieved, when the low frequency oscillator has a ramp up (attack) or ramp down (decay) output available.This is then used to modulate the frequency of the low frequency oscillator, speeding up or slowing down the oscillations, hence having spin and spiral features. The combination of these motions and translations can result in *bidirectional* or *multi-directional* motions. *Contraction* for instance, an element of *multi-directional* motions, indicates that a motion comes from outside, and retracts inwards. This can be simulated in a four-channel setup, in that sound pressure level is used, alongside dry to wet signal ratio of reflections, making it appear as if a sound is far away, quiet and diffused, which then comes closer, “contracts" towards the listener and becoming louder and less diffused from multiple (circular) directions. *Dilation* is the reverse of this process.

In moving sound objects, speed plays a key role in how we perceive the sound. Generally, it can be said that the slower a sound source moves, the better the ability to locate the sound specifically, and analyze its features. If the speed of circular movement increases to certain rates, sounds become perceived as oscillating between the sides of the head, left and right. When the speed is increased even further reaching audio rates, a so-called "*persistence*" or in other words inertia, happens. This is the phenomenon where a single sound source extends itself through space (and time), to be perceived as a single, diffused, spatial auditory event (Blauert, 1997, p. 47). This means that at high speeds, a sound would exist at all points of the trajectory at the same time.

Acousmatic Sound and the Spatial Narrative

Pythagoras, the renowned Greek mathematician and philosopher, first gave an example of acousmatic sound when he required that his students listen to him from behind a curtain, so as to focus solely on his teachings and not the person providing them. Acousmatic sound is sound, the source of which is unknown to the perceiver. Any sound without a physical reference is acousmatic. In this case, we can recall Pierre Schaeffer's use of the word. Acousmatic sound was used by Schaeffer to describe the well-known "Musique Concréte ". Usually, almost all sounds projected from a loudspeaker can be acousmatic, because even if the sound is classical music, the source is not perceived and the perceiver is not sitting in front of the ensemble playing it. Therefore the source of actions and gestures of which produce the music differ with electroacoustics. Emmerson (1986) argued that regular attendants of concerts with live performances can have a feeling of disorientation when they attend a loudspeaker concert in which classical music is played through loudspeakers, as they are used to seeing the source of the sound perceived. With the use of loudspeakers in daily life, the acousmatic sense of sounds might happen to disappear, as loudspeakers themselves become the source and event behind sounds. We do not always see a cat when we hear it through a loudspeaker knowing where the loudspeaker is, although the origin of this sound is actually a physical animal.

Smalley (1997) explains *function attribution*, which is an ongoing operation that takes place in the unconscious, where, while a sound is being perceived, the cognition is constantly analyzing what meaning and which attribute to give the sound during its development and subsequent dissipation. *Function attribution* is also an operation which is subject to changes throughout the development process, and is not decided or fixed, as this is an operation unfolding through time, and events can change unexpectedly. Therefore, anticipation, as well as the unfolding of events, both decide *function attribution*. Furthermore, it can have several connotations, which is why context plays a defining role in the acousmatic narrative. Two sounds moving symmetrically, in opposition can either be perceived as separation and alienation of each other, or can be perceived as a choreography of harmony. As soon as one sound changes behavior, anticipation is broken, and the "in the now" experience of what unfolds is the process of finding out and judging the sonic qualities.

In the musical discourse, a note can serve as the tiniest component of a musical unit. The combinations of notes become greater classifications, like leitmotifs, which in turn make up greater segments, these segments comprising a piece themselves. As for spatial composition, the single smallest, and easiest point or movement to be localized in space makes up the smallest spatial unit. A linear ascending movement between two points, a straight line, is the lowest unit of that order. Duration also plays a role in the perception of single units. The shorter the perceivable duration is, the lower the unit.

Batchelor (2015) simplifies this in writing: "composers connect and combine these sound images, constructing (loose) narratives by building and manipulating perceived relationships between them". Denis Smalley writes:

> The acousmatic experience of a given landscape builds spatially as we aurally explore its terrain, yielding a complete 'image', which exists outside the bounds of time. Time effectively becomes space. For the listener, this image therefore persists as something akin to the memory of a photographic still – an encapsulation of the experience in a single moment. It is the combination of a series of 'stills' that together form the narrative experience of a work. (Smalley, 2007, p: 37–8)

There is no single meaning to a moving sound or space. It is rather solely dependent on the context in which this phenomenon takes place. A movement of two sounds moving in symmetrical, opposite directions can have two or more individual connotations. One meaning could indicate synchronicity, while another can be of opposing nature. This movement then only takes on a meaning when it is combined with the events preceding and following. If in the example above, one sound stops while the other keeps its movement, another meaning, that which is perceived as synchronicity, is broken. If the sound that has been halted disappears, the movement would take on yet another meaning, which is "vanishment" or dissipation. This combination of single events can be called "*Chunking*". *Chunking* is a term also introduced as the group of events that happen in a period of 0.5 to 5 seconds, which contribute to the holistic presentation of a sonic event. "Quantal Elements in Music Perception" helps identify important features and parts of a body of sound or music, which make it the way it exists in the perception. As sound unfolds through time, a sequential arrangement might contribute to the perception of the greater sound phenomenon. Rolf Inge Godøy mentions that: "(...)it is also well known from psychoacoustic and music perception research that sequentially occurring elements contribute to holistic, and in a sense 'atemporal', perception of features(…)" (Bader, 2013, p. 113). So does pitch, timbre, and other elements including motion, contribute to the end product holistically. Just as single elements of sound contribute to chunks of up to 5 seconds, so do these chunks contribute to the piece of composition. This is also a major factor in motion perception: in order to perceive the motion from point A to point B, the perception of both points should be assured.

At a larger-scale, the spectrum can be divided into three other categories: *micro*, on the level of a single sonic event e.g clap, where microscopic fluctuations and modulations of dynamics, pitch, and time occur. *Meso*, a level higher than micro, normally corresponds to the above mentioned *chunk*, holds leitmotifs, melody, rhythms, textures and so on. And finally, *macro* level is made up of multiple *meso* chunks. Rolf writes:

> This is why Pierre Schaeffer stated that for any sonic object we always have the duality of a larger scale context of the sonic object and an internal contexture of the sonic object, as well as "the two infinities" of music, meaning that in listening to musical sound we can intentionally zoom in and out, similar to Husserl's idea of being able to zoom in and out of any musical excerpt in our memory" (Bader, 2013, p. 117)

The comparison between film editing and sound narration in sceneries has also been made by Katharine Norman (1996), who suggests that there exists meta-information carried not by the images themselves, but by the combination and sequence that these images take.

Sonic events are contextualized in the perception, independent of their neutrality. Space communicates. An installation in a room with hidden loudspeakers works to hide the sound sources as perfectly as possible, although the signs and staff leading to this room of exhibition, as well as the building and eventually the city and the country communicate information to the perceiver. The artists themselves and their portfolio is enough information to generate anticipation.

Brech & Paland (2015) argue, similarly to Smalley (2007), that by creating a sequence of instances that relate to each other, these instances can be grouped into an anecdote. Brech & Paland (2015) write quoting Simon Emmerson: "(...)the listener may create a series of »moments« that might group themselves into a possible narrative. These moments may be constructed (…) through events articulated by change in the sound – our attention is preferentially drawn to change". Change indicates direction, as point A changes into B. This change in sound can also be attributed, whether simultaneously or not, to its physical position and can be a manifestation of a combination of many aspects. Trevor Wishart's Vox 5, composed in 1986, is a wonderful example of how voices turn into sounds that indeed unfurl into a sequence of events to the perceiver, an indication of spectro-morphological evolution. Pierre Schaeffer, as mentioned before, changed the sequence of notes using knobs and faders that controlled the playable turntables, determining the sequence of notes and hence the mood or "direction" of the "Sentence". Similarly, changing the relationship between the two spatial images might generate a perceptual change in mood, as would a change of chord in music. Emmerson (1986) differentiates between two types of mimetic images, in which we perceive sounds. The first is the immediate "*Timbre*", where the second is "*Syntactic*". To explain the first, it is simply the resemblance of a pitch of a sound to that of an object that is familiar to us. As an example, a high frequency transient sound which is located above might be perceived (through functional attribution) as a bird. *Syntactic* is, on the other hand, the imitation of a sequence of events that occur in a natural manner. If tires of a car screeching are heard, a car crash after it might be anticipated. This can be to the benefit of a composer or sound designer, becoming a manipulative tool based on expectation. Sound organization and the ability to communicate concepts sonically are cross-cultural. This kind of art abides by the "Universal Law ", as referred to by Lévi-Strauss (Emmerson, 1986, p. 21).

Frequency (Pitch), Timbre and Space

Sounds are laid on a continuum from note, through node, into noise, where note and noise are opposite extremes, according to Smalley. Note is divided, according to Denis Smalley (Emmerson, 1986), into three categories: *note proper, harmonic spectrum, inharmonic spectrum* (see Figure 11).

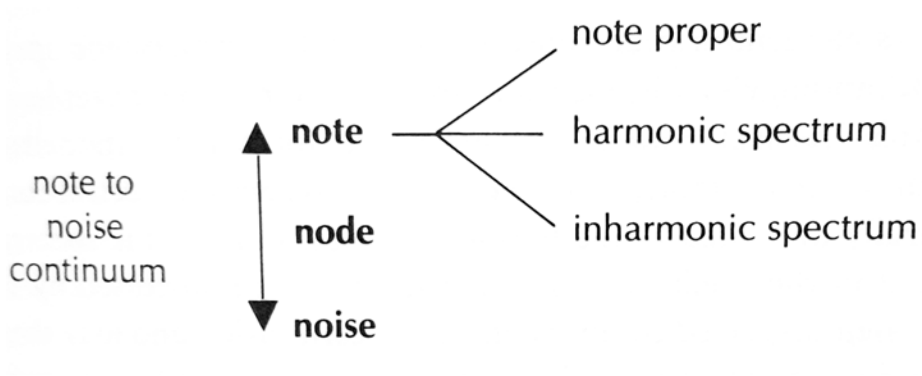

Figure 11. Emmerson, S. (1986). Spectral Typology.

Note proper makes up the absolute pitch of a sound with its intervals and combinations. This is a cardinal part of music theory. *Harmonic* and *inharmonic spectrums* lend sounds an identity. They encompass the fluctuations of overtones and resonances in a spectrum. Harmonic partials act in a coordinated and organized manner, in contrast to inharmonic partials. An example is the sonic difference between a flute and a violin. The two instruments can emit the same note proper, but vary in the harmonic and inharmonic spectral structure, hence emitting different timbres. The *inharmonic spectrum* encompasses partials which are not of the archetypal structure seen in *harmonic spectra*. *Noise* is the opposite of a *note proper*. It is a sound whose pitch is *not* clearly perceivable. This is due to fast and aperiodic fluctuations happening in a noise spectrum with little regularity. Everything that falls in between note and noise is a combination of both, thereby named *node* or *nodal spectrum.*

Natural sounds don't have a steady spectrum, even more so when motion is present. The change in spectrum might increase or decrease depending on the environment and the medium in which sound travels. Doppler's effect not only induces change in the spectrum because of velocity, but also bends the note proper. This is, in this case, proof of how drastically physicality and materialism can restrict and intervene in the perception of things. Indeed, these three categories can be controlled in a Eurorack environment, molded and bent to fit a specific use or number of uses. For example, to change the harmonic and inharmonic structure of a sound while maintaining the pitch, a filter or phaser can be used to roll-alter partial structures or to restrict them to certain bandwidths. In additive synthesis, partials (harmonics) emerge through superimposing multiple waveforms. In additive synthesis, the number of waveforms superimposed, spectral position of the partials, the number of partials, the amplitude of certain partials, and the distance between them (also called spreading) can be controlled. All these parameters are independent of the pitch, but sometimes induce drastic change into the nature of the sound. Furthermore, in the nodal spectrum, the difference in relationships between the three pitch categories mentioned above make up the tonal qualities of sounds. The spectrum of a cymbal (nodal spectrum) for instance has a bigger inharmonic to harmonic ratio than that of a piano, as well as having a higher spectral density. Sounds with such characteristics fall under the category of *pitch-effluvium.* It is a result of such high density spectrum structure that exhausts the ear in trying to perceive a pitch (Emmerson, 1986). That is why it is easier to identify the pitch of a piano string than to identify a cymbal. Having pure tone oscillators as well as noise generators in modular synthesis lends freedom to the sound designer, as to how much of what makes a spectrum and the purpose of it. Aside from having the possibility to input any kind of signal from external devices, such as computers, into synthesizers, it is possible to achieve a noise like sound only by modulating a wave heavily e.g. high index, audio rate frequency modulation. Inversely, a sine wave can be achieved by filtering noise (subtractive synthesis). The ability to control the depth of modulation or the frequency and bandwidth of filters allows to incrementally control the characteristics of the desired sound.

Spatial attribution to pitch is not a new phenomenon. Qualities can be attributed to different sounds, like big, little, dull and sharp. These are attributes that have a physical approach, or be suited for describing physical objects. This provides a supporting ground that sounds are perceived not as mere vibrations, but also as vehicles that carry physical and spatial (three dimensional) qualities. Pitch is understood as the number of oscillations per time unit, where the shape of the cycle changes the timbral qualities of the sound (Bader, 2013, p. 79). Emmerson (1986) describes the morphological approach of sound as having multiple attacks that make up the pitch, called *The Attack-Effluvium Continuum* (see Figure 12).

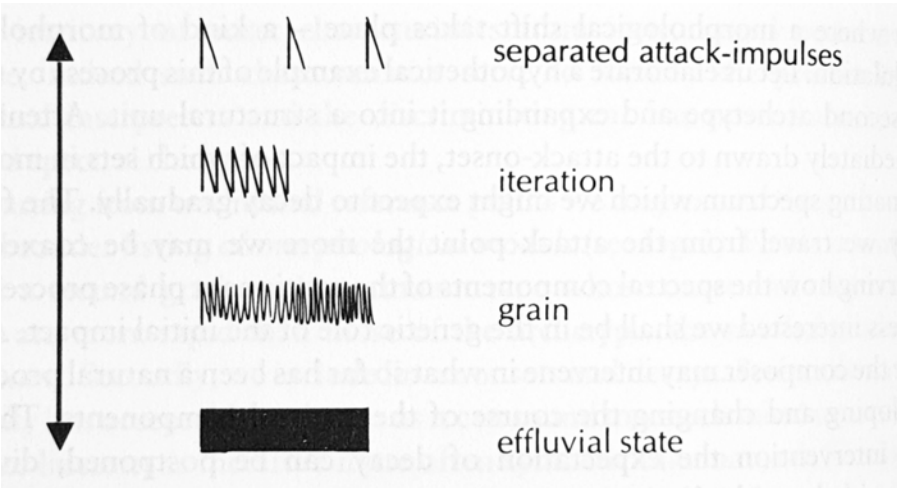

Figure 12. Emmerson, S. (1986). The Attack-Effluvium Continuum.

Pitch-effluvium states that the compression of impulses (or other waveshapes) makes up a pitch. In *attack-effluvium,* the compression of multiple attacks (impulse) together at a certain speed makes up an effluvial *state*. Starting from separate *attack-impulses*, which are perceived single impulses, *iteration* is the first stage of the merging of the attacks, so that they gain a quality as a whole. The boundaries of perceiving them separately start to vanish. Merging them further together, the grain stage is entered, where individual impulses surrender all qualities that help distinguish them from the whole. *Effluvial state* is reached after even further compression of impulses together, so that the whole of the sound loses the last of the granular qualities and it becomes perceptually easier to follow the general envelope of the sound, and not its single parts. Texture and timbre are determined as a result of different shapes of attacks, like *attack-impulse* (instantaneous and impulse shaped), *closed attack-decay* with fast decaying resonance (like gong impact sounds), *open attack-decay*, where the resonance is not sustained, but is longer lasting than a closed attack-decay archetype, and finally *graduated-continuant*, which can be categorized as a gradual sustain (see Figure 13).

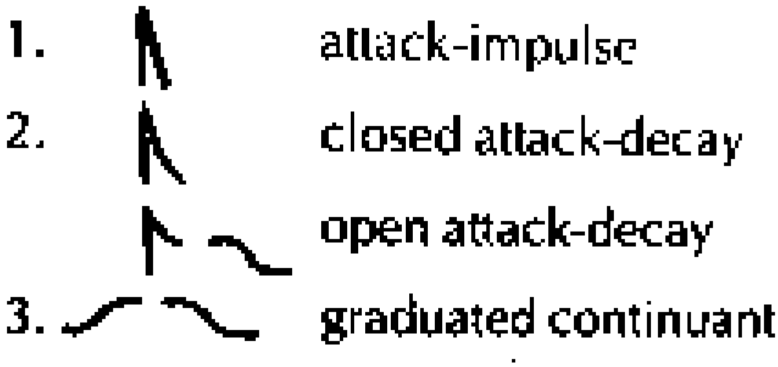

Figure 13. Emmerson, S. (1986). Morphological Archetypes.

Furthermore, both attack-decay archetypes (closed and opened) can be reversed. Decaying must not always be the disappearance of sound. Defined as the termination stage of sounds, decay can increase in intensity for instance. In addition to reversed impulses, *linear attack-decay*, *linear and swelled graduated continuant,* with their respective reversed forms (see Figure 14) make up tones that do not stand alone, but are merged together, like a chain.

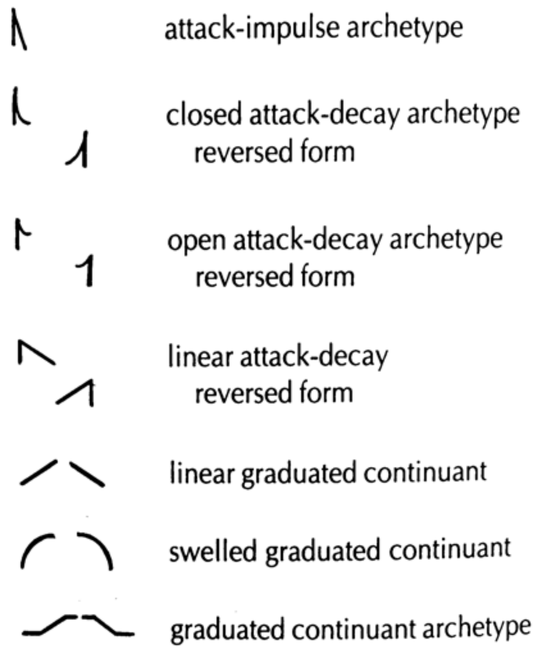

Figure 14. Emmerson, S. (1986). Morphological Models.

In figure 15, *morphological stringing* is explained as the merging of single impulses. *Open continuant phases* correspond to *open* and *closed attack-decay*. *Merged correspondences through crossfading* are defined as gradual instruction and reduction between impulses. Lastly, different combinations of different impulse shapes generate new forms, especially when reversed, representing the third morphological stringing, *reversed onset-terminations leading to new onsets.*

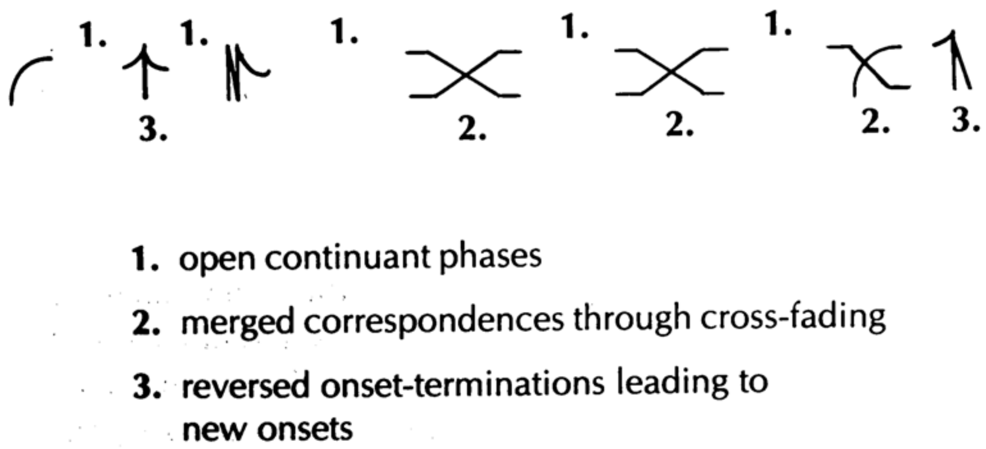

Figure 15. Emmerson, S. (1986). Morphological Stringing.

A pure tone of a certain pitch, like a sinus tone, has a strict periodicity between gradual cycles, and the shape of the single oscillation has minimal deviations in time and spectrum. Natural sounds, or broader, complex sounds do not have a constant type or shape of oscillations through time and spectrum. Helmholz claims that sound color depends on the microstructure of the vibrating object emanating from the object, adding that most complex sounds can be re-synthesized by superimposing multiple sinusoidals. This approach is known from Fourier's theorem, which is "analysis by synthesis", in which the ear breaks down a complex sound into simpler parts. Additive synthesis touches on this idea, where sound color and other sound qualities are determined by the sum of the different parts of a sound.

Stumpf argued that sounds may take up three different attributes: height, intensity, and volume (or extention). High means in the literal sense, where the sound is positioned on the vertical scale, depending on its frequency. Intensity is defined by the magnitude of the oscillations, and volume is how much space the sound occupies. However, Stumpf separated between the above-mentioned attributes, claiming that height and intensity are objective as they can be measured. Volume, or extension, is subjective, however. (Bader, 2013, p. 82). As an example, low pitched tones with poor fundamental vibrations tend to be perceived as large or occupying a greater volume, whereas the opposite applies to high pitched sounds. Frequency as well as intensity are considered physical attributes since they can be measured, while volume or brightness of sound tends to be subjective and exists merely in the perception. Stumpf also claims that any sound, complex or pure, can be assigned a spatial quality, as almost all sounds have pitch and intensity by nature. Physically, pitch and intensity can be assigned a spatial "dimensional" attribute, of which they are integral, whereas in volume and sharpness, the spatial attribute is distinguishable from them and not fundamental to them. A three dimensional temporal envelope of a complex sound can be achieved using two characteristics of sound color, namely "inner" and "outer" moments. While both of them make up the whole of a complex sound, they are indeed different when inspected in isolation. Inner moments define the structure of the partials at a given time, along with their respective amplitudes, and can be defined by a temporal envelope. Outer moments comprise the kinetic envelope of the sound over time, including onset, decay time, transient, noise, modulations and so on. Inner moments can determine distance depending on low frequency intensity, while outer moments determine distance depending on the sound's intensity. Furthermore, Stumpf also found that the general structure of the sound is extremely important to its recognition, similar to Denis Smalley's Spectro-morphology and to the work of Rolf Bader (2013). He experimented with experienced musicians, letting them listen to instrumental sounds of which the onset and decay were eliminated, leaving only the sustain stage. The listeners found it hard to identify what the sounds were (or from which instrument they were emanating) (Bader, 2013, p. 84). This is proof that increments of the whole of a sound are crucial to its identification.

Arnold Schoenberg argues that there are two (or more) dimensions to spatiality in music, relating to pitch. The vertical dimension refers to the pitch-height concept, and the other dimension refers to the linearity of rhythmical elements in time. According to Schoenberg, a chord can be viewed as vertical notes stacked above each other, sounding simultaneously (traveling on the horizontal linearity of time). It is rather the relativity and mutuality of sounds to one another that defines the perceived space. Distance, which is a description of space between the notes that make up a chord, is the mutual relationship between the notes. The transposition of notes yields the same "feel".

Xenakis mentions that these models of pitch and space indeed exist, however only existing in the human comprehension and nowhere else, rendering them counterfeit. This is only a construction of perception, a reference created by humans as part of notation systems, in order to aid imagining and remembering (Harley, 1994).

Glide, or slewed pitch, is the gradual or continuous change of pitch or frequency, giving the change in pitch a fluid-like, bent feel in contrast to edgy and abrupt pitch change. This feature of pitch-change carries hints about the sound object, and the manner in which its character behaves. Slewed pitch might seem softer due to its non abrupt nature, while it can also induce a feeling of drunkenness and heaviness, as a characteristic of the sound object. Control over slew time, which is the time needed for a pitch to gradually change to another, can be modulated in a synthesizer environment.

Pitch-Height Effect

Carroll C. Pratt (1929) has proven in early experiments that pitch takes on a spatial character, a claim and examination made by Carl Stumpf (Stumpf, 1883). The usual association of high frequencies is with high positioning, while that of the lower frequencies is with lower ones. An experiment conducted by Carroll C. Pratt, (1929) proved the results by asking subjects to locate the position of sounds ranging from 256 Hz to 4096 Hz, increasing the frequency by an octave upwards. The pitch-height is also notable in many designs, like loudspeaker designs, where the tweeter almost always sits on top of the woofer in two or three way speakers (excluding coaxial designs). Moreover, birds naturally soar above the earth and mostly have high pitched voices, while the earth beneath us radiates rather low frequency sounds like those of tectonic plate movements. From a historical and physical point of view, this pitch-height scenario is embedded in our understanding of the world. Another probable explanation to the pitch-height effect, according to Tilley & Cabrera (2003) is the fact that the coloration of sound sources located above the head level tend to have interaural colorations, from the highest audible frequencies to the low 700 Hz (while other research claims even lower frequencies). Colorations due to ground and torso reflections tend to interfere with the direct sound waves, even at the lower end of the spectrum, which causes modifications in the spectrum at the ear level when an elevated sound is at hand (coloration differs between people, because of the differences in body shape and outer ear).

Height Phenomenon

Humans perceive sounds on both median and horizontal planes. This can be traced back to several factors. Sounds tend to be perceived with an increased elevation angle when an identical stereophonic signal is played back with increasing angles between the two horizontally placed loudspeakers from 0 to 240 degrees, the frontal center loudspeaker being at 0 degrees (Lee, 2017). In particular, angles from 0 to 120 degrees had a frontal or elevated frontal position, while speakers with base angle between 180 and 240 had an elevated back image. Between a 300 and 360 degrees base angle, results yielded back image positions. The most dominant feature of the phantom image elevation effect is nevertheless the spectral make up of a sound. For instance, in an experiment conducted to investigate elevation effect, Lee (2017) found that sounds with less high frequency energy in comparison to the lower frequencies tended to occupy a lower degree of elevation than sounds that possessed more energy in the higher frequency spectrum than the lower ones. This is precisely when the base angle of the second loudspeaker was located at 120, 180, and 240 degrees. It is also assumed that the ability to differentiate between front and back is heavily dependent on the presence of midrange frequency bands. Measurements using HRTFs have yielded results that can be linked to the spectral cues of elevation, as well as front and back differentiation. The so-called *directional bands* contain cues that help perceive an elevation image. Bands around 8kHz have an increase in dynamics with stereophonic reproduction, in contrast to monophonic reproduction, within all base angle positions with varying ranges. This hints that these bands contribute to the elevation effect. Furthermore, it is suggested that the presence of the bands at around 500 Hz and 4 kHz are linked to frontal sound source perception, while bands at around 1 kHz are linked to back sound source perception (Blauert, 1997). It is also speculated that the height phenomenon resets at around 1 kHz. This means that from around 60 Hz up until before 1 kHz, the height phenomenon is somewhat linear and consistent. At 1 kHz, the height of the perceived bands drops again to ear level, showing consistency with frequency bands around 60 Hz, hence hinting at the non-linearity of the pitch-height effect over the frequency spectrum. This might also be an indication of the front-back confusion that subjects in different tests presented and is indicative of Blauert's theory of *directional bands* (Lee, 2016). There are further suggestions that the torso reflections of sounds coming from above provide important cues in the perception of elevation for frequencies up to 3 kHz. Allegedly, the greater the elevation angle (if the center loudspeaker in the Lee (2017) experiment is elevated to up to 90 degrees, meaning directly above), the greater the interaural delay becomes. Moreover, Lee (2017) hypothesizes that in addition to the above mentioned cues, the interaural crosstalk delay generated by stereophonic horizontal loudspeakers outputting the same sound at the same time must be considered, which is equivalent to the interaural time difference generated by one of the two loudspeakers that contributes to the elevation effect. Therefore, the perception of the interaural crosstalk delay might be perceived as torso reflections, although this needs further investigation. Localisation is improved when the sound is familiar or known to the observer, since differences in frequency fluctuations and linear distortions are better perceivable, when the observer has prior experience (Blauert, 1969). Certainly, there is some degree of individuality in directional frequency bands (Itoh et al., 2007). Other early research supports that elevation, from the view point of interaural time difference, relies on being on a vertical axis inline with the head (top of head to neck). The interaural time difference on this axis is always the same, if the sound source is on this certain axis and thus, in order to better locate a sound source, subjects tend to introduce head movements in order to change the interaural time and level difference, as well as the spectral make up at the pinna (de Boer, 1947).

Waveforms containing different levels or dynamics of different frequency bands can be interpreted as spectral cues, depending on the loudspeaker placement or phantom image positions. Even if not specifically for the purpose of elevation effects, these modulations might cause sound sources to appear located in the median plane, or passing through it. The quadrilateral loudspeaker setups tried for the thesis contain different loudspeaker angles. Assuming that a subject is standing in the middle, different angles to different loudspeakers arise, including some angles mentioned by Lee (2017, 2016), and when playing back spectrums that involve Blauert's *directional bands* (Blauert, 1997), the height dimension can be simulated and perceived psychoacoustically. When subjects move around, the degrees of the loudspeakers change but as long as the subjects stay in one of the possible loudspeaker pair axes, the effect can be perceived. Furthermore, in the four-channel setup, it is possible to share a single signal over four loudspeakers, creating combinations of pairs that might induce this effect (diagonally for instance). A sound that has more energy in the 8 kHz region might be perceived as above or higher than sounds that lack energy in that domain. Reverberation, despite its diffusion and broadening of the image (which contributes to less localization ability), can contain a variety of bands (dependent on the reverberation type and settings). This increases immersion further via spanning a vertical plane and connecting it seamlessly with the horizontal one.

Panning a single sound source in a four-channel format would require the distribution of the same sound signal to all four channels, while controlling single loudspeaker dynamics simultaneously (also with short delays) might also contribute to the interaural crosstalk delay, which is equivalent to the torso reflection interaural delay. Therefore, it can be assumed that the elevation effect can be targeted in a four-channel format as an artistic and a controlled parameter, whether in the disposal of certain frequency bands responsible for the spectral cues (like increasing or decreasing energy of directional bands, filtering bands of around 1 kHz and so on), or in the utilization of different loudspeaker pairs that contribute to interaural crosstalk delay at will. The so-called sweet spot is affected by the movement of the subjects, not to mention the effect that the head movement of the subject alone displays when stimuli have a duration of 1 second, in comparison to bursts (Itoh et al., 2007). Height is dependent on the Frequency response of sound (physical), and also related to pitch-height effect (cognitive or psychoacoustical). With experimentation in four-channel quadrilateral format, height effect was achieved without height loudspeaker placement.

Synthesizer Architecture

In the following section, the different elements of a synthesizer will be defined and brief examples will be given of their functions in relation to the four-channel installation presented in this paper.

Interface

Interfacing with computers and loudspeakers not only reduces the production costs by reducing the number of modules necessary for producing certain effects, but also makes the transmission of signals between modular synthesizers and computers easier. In its essence, it is the embedding of a computer into the synthesizer as the variations in methods of connectivity rises, thereby making the computer a part of the synthesizer, and changing analogue signals to digital ones and vice versa.

Utility

Utilities are usually multi-function modules. They serve as small computers or doors, which at defined times let a signal through. Stockhausen's Sound Mill uses a similar system such as a switch. It routes one signal to more than one output, or multiple inputs to an output. The signal flow can be controlled either manually or using control voltage to select which input or output is active. This can be used to route audio signals to different loudspeakers, as well as gates (a gate is a sudden high voltage, used as a trigger or modulation source. It can either be on or off, and has no levels in between), control voltage, and anything else that can be outputted from the synthesizer, to any destination. Logic is another form of utility modules. Boolean logic (stemming from Boolean Algebra in mathematics) is a set of mathematical operations that can be calculated, resulting in a mathematical combination of signals. In their essence, logic modules are computers that perform certain operations without the need of human attention.

A signal available at the input of a sample and hold module is be sampled and outputted, otherwise being held, as soon as a trigger or gate is received at the clock or trigger input. Slew rate limiters are processing units that change a stepped signal into a gradually changing signal. In doing so, it restricts the number of volts outputted to a certain duration, so that the number of volts at a given time does not exceed a given threshold. If a sudden change happens in the signal, a slew limiter slows this change down, depending on the rise and fall settings (corresponding to rising time and falling time of control voltage). This can be used for pitch, to add glide and portamento, smoothing out edgy control voltages.

Precision adders are modules that combine control voltage, keeping pitch intact if pitch control voltage (voltage per octave) is being calculated.

Attenuators amplify or weaken an incoming signal, while attenuverters can perform the same functions, in addition to inverting the signal (turning positive voltages into negative ones). The amount of attenuation or attenuversion can either be modulated with a control voltage, or adjusted manually.

Clocks and Triggers

Clocks and trigger modules are devices that have an internal sample rate that measure time and generate an event (high voltage) at a given duration. Clocks mainly control the speed of triggered events that happen in a synthesizer. When controlling the amplification of an audio event (using an attack-delay envelope), pushing the clock that triggers the envelope to audio rates simulates *attack-effluvium*. Clocks can have modulation inputs where control voltages can be fed into the module to control clock rate. A Turing machine is a combination of an organized and random clocking, with voltage outputs. This encompasses Alan Turing´s, Markov Chain´s, and Xenakis's stochastic mathematics. It lends control over probability of events happening. Its function is outputting organized yet random gates and voltages, with a loop function. If an intriguing random sequence of notes and triggers is sampled, looping this sequence would then make it repeatable, while manipulating other aspects of the sequence like root note (when working with one volts per octave) or the maximum distance between the highest and lowest voltages, etc... Some Turing machines also have modulation inputs where voltages can be fed in, in order to change aspects like trigger sequences, timing, time bias, and looping of trigger sequences, as well as its slow or fast morphing into random or steady clocking. Turing machines serve as the centerpiece of generative sound, as it is possible to output random sequences while restricting them to a set of rules, making them both organized and random.

Modulators

ADSR modules are envelope modulation sources that have a four stage progression. A trigger is needed to trigger the envelope. Some envelopes, however, have a loop function and thus act as low frequency oscillators with four stages. Each stage can be modulated with a control voltage or set manually. Attack, decay, and release correspond to timing, as in the duration it takes for an attack stage to be completed, while sustain corresponds to voltage strength output, as long as the gate of the input trigger is held at a high level (length of gate equals the duration of the sustain). Depending on the gate length in the trigger input of the envelope, the sustain level stays active, before the envelope goes into release, which is triggered as soon as the input trigger signal disappears.

Envelope generators are similar to ADSR modules in that they output envelopes, but they usually have fewer stages. Typically, two stages are available, attack- and decay-make, for envelope generators. Some envelope generators can be set to hold and sustain a certain voltage as long as the trigger signal at the input is high.

LFO is short for low frequency oscillators. They are, as implied, oscillators which generate waveforms slow enough that make them suitable as a modulation source. Some LFOs can output different waveforms, as well as work in quadrature (90 degrees phase between all outputs) phase with variable phase shift, or in divisions of the input reset (clock) signal. Typical waves are sine, triangle, ramp, and square waves.

Joysticks add a spatial and gestural dimension to sound in general. Having a joystick in the Eurorack environment translates hand movements and gestures into control voltage, which can be used to position a sound over a quadrilateral four-channel setup. This makes the gestural movement of sounds, but also the modulation of location with spectral change as well as other modulations in the system possible. Furthermore, a joystick can control position, pitch, texture and/or time, either each factor alone or several factors simultaneously. When a clock is modulated by an axis output of a joystick, the result is a spatial gesture that translates to a change in the speed of events. Space in this case, determines time, which in turn determines spatial movement of sound. This is similar to the concept that John Chowning designed in 1972, making spatialisation based on gestures that composers themselves generate (McGee, 2009).

Mixers and Voltage Controlled Amplifiers

Voltage controlled amplifiers (VCAs) are one of the main elements which contribute to the spatialisation of sound in a channel-based format. Amplitude panning in a four-channel setup is possible, using simple triggers with envelopes. The ability to control amplification, timing of amplification, and amount of amplification grants control to spatial attributes, like position, movement, and depth.

Other panning modules have multiple VCAs built in and can perform multiple channel vector-mixing, -panning or -synthesis. Simply put, two bipolar signal inputs, for instance, control the positioning on both X and Y axes respectively. While this is similar in function to having four channel amplifiers (using four different modulation sources to amplify), using two bipolar signals to control position grants control over the axes which promotes experimentation spatially.

Filters

Filters manipulate spectral content. They act as a window (or barrier) for certain frequencies based on the frequency cut off position. A number of spatial effects can be achieved through filters, when working with multichannel formats. For example, using filters (or resonators), spectral space (where every loudspeaker receives a designated filter frequency), the height phenomenon (through amplifying directional bands), width and depth or spatial position (through emphasizing certain bands) can be achieved in a four-channel format.

Oscillators and Sound Sources

Oscillators are wave generators that output a waveform, and are used as a sound source in synthesizers. Some oscillators work on the basis of additive synthesis where a final waveform is outputted as a result of adding different waveforms, while other oscillators work on the basis of subtractive synthesis, where the removal of certain bands and the use of certain techniques yields a desired waveform. Control over voltage per octave, pulse-width, symmetry, warping, folding, and waveshaping is possible manually or using voltage at the corresponding inputs.

Effects

Effects in a synthesizer usually range from reverberation, echo or delay, to spectral manipulation, made possible by equalizers, waveshapers, bit crushers, flangers, chorus, phasers, distortion, vocoders, granular and spectral processors. Effects change the spectral content of sounds. Reverberation and delay, on the other hand, are rather direct spatial manipulations (like delaying a sound over a loudspeaker setup, or the illumination of space using reverberation).

Space as a controllable Parameter in Eurorack

Object-Spatial Character

Sound Pressure Level Character

In defining the dynamics envelope, it is necessary to explain its three main stages. The first is attack (onset), which is the first instance that a sound emerges until it reaches its peak. A peak can be sustained or directly fall into the third phase of decaying. Bader (2013) explains three different dynamics types: *impulsive*, *iterative*, and *sustained*. These dynamics are intertwined, for sustained dynamics are made of very fast impulsive units. This is analogous to pitching down a sound until the single oscillations are heard as clicks. The pitch-effluvium by Denis Smalley is another explanation to the concept of impulses-sound spectrum. In volts, an envelope starts with 0 volts, goes through a rise to +10 volts, might be held at 10 volts for a definite time, and then fades out to 0 volts.

The nature of sounds and their spectral dynamics envelope reveal change in the spectrum, and can include their position, motion, and the space they appear in. Such envelopes communicate, for instance, information about the event that caused a sound, much like the spectral dynamics envelope of a gesture of a musician playing the violin. The concreteness of events bundled with the dynamic sonic representation of the sounds fuse together. Yet, the ability to manipulate sound grants us the ability to manipulate the mental representation of objects or events that caused it. Spatial information is also a key point in dynamics. As a sound gets further from the listener, its overall dynamics (intensity) decreases. The dynamics of the spectral content, like the lower frequencies, also decreases with increasing distance from the listener. A ramp-down envelope on the global dynamics of the sound, as well as on the roll off of some of the low frequencies, simulates a similar effect.

This is not to discard other elements that interact when an object is in movement, such as the Doppler effect and other variables like reflections, etc... Denis Smalley argues that: "The dynamic profile articulates spectral change: spectral content responds to dynamic forces, and conversely, dynamic forces are deduced from spectral change". He goes on to elaborate:

> In the environment, when a sound approaches the listener its spectral and dynamic intensity increase at a rate proportional to perceived velocity. Moreover, the increase in spectral intensity permits the revelation of internal spectral detail as a function of spatial proximity. (Emmerson, 1986, p. 68)

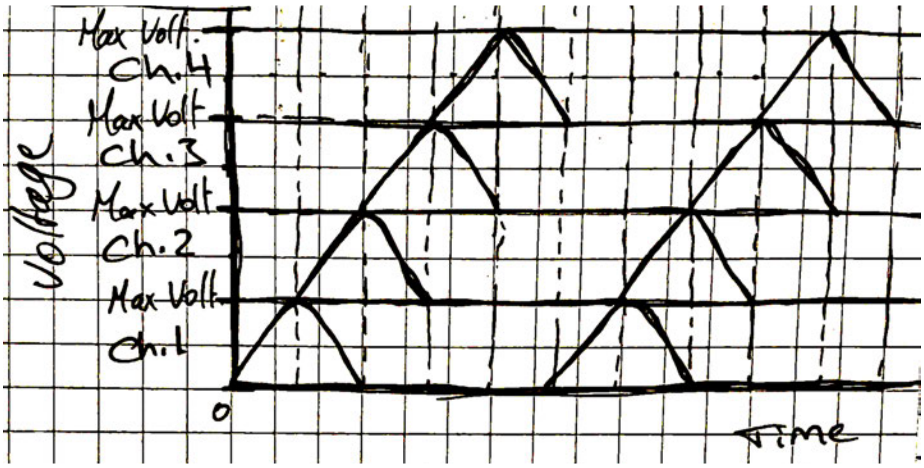

Figure 16. Sketch of Quadrature LFO (Triangle waveforms) Simulating Circular motion (clockwise)

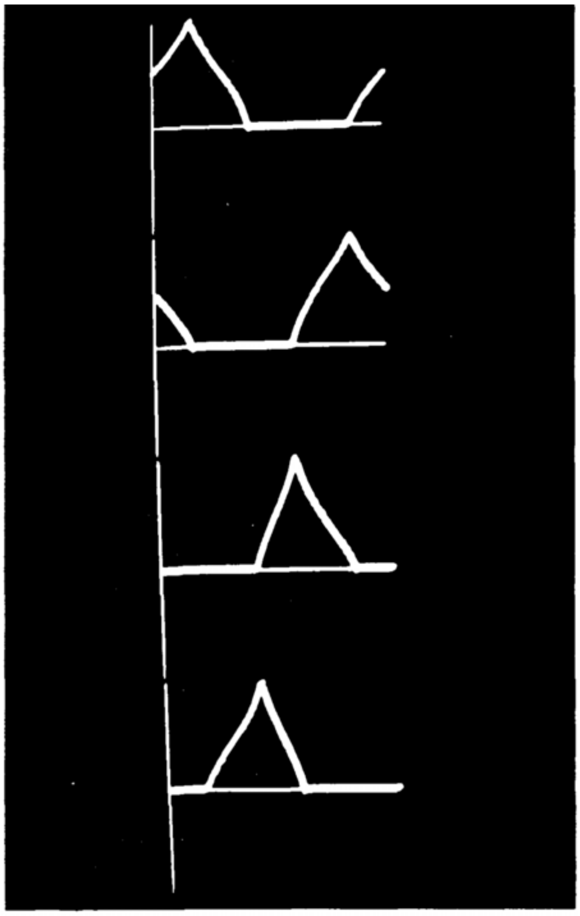

Figure 17. Chowning, J. (1971). Control functions for Angular Displacement.

John Chowning's control functions for angular displacement representation (figure 17) is a visualization of a circular moving sound source. Similarly, sending phased triangle waves into four different amplifiers' voltage inputs, in a four-channel format, resolves into a circular movement, similar to Chownings.

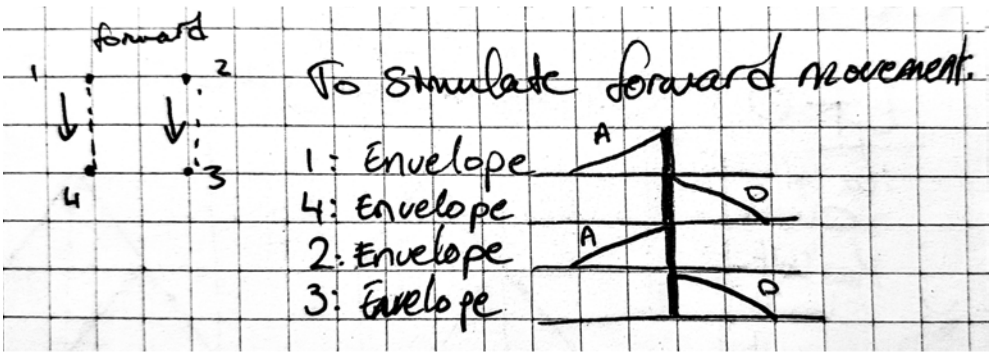

Figure 18. Sketch of Forward Motion Simulation using Attack-Decay Envelopes *Note:* on the left is the loudspeaker setup of the total number of channels, and the defined forward direction.

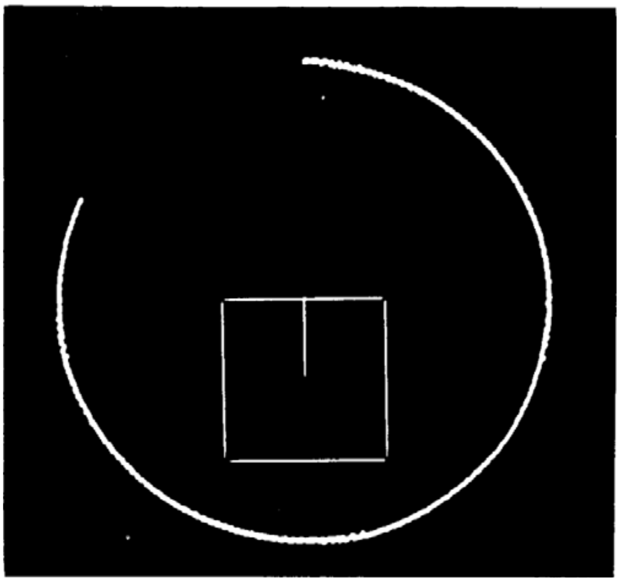

Figure 19. Chowning, J. (1971). Sound Path of Moving Source (Clockwise) around Listener Space.

Additionally, the dynamics envelope articulates time. An impulse (transient-like) dynamics envelope simulates the natural passing of time. Consider a sound that begins abruptly (short attack duration), initiating the acoustic phenomenon until the peak intensity is reached, and then fades away(decay duration longer than attack duration). Using this logic, it is therefore possible to reverse events. An envelope that has a longer *attack* duration than decay duration might indicate the reversal of the event or of an impulse. E.g. rhythmical hi-hats can be simulated in a Eurorack environment using white noise. With an exponential Ramp down envelope, it is possible to approach a realistic hi-hat sound. If and when this dynamics envelope is reversed, the sound of reversed hi-hats is simulated, as if played backwards. This differs from sounds played in reverse, however. Here, the same spectrum is let through, only with different intensity envelopes that simulate the opposite of how something naturally sounds. Therefore, the shape of the dynamics envelope can have an effect on the perception of time direction.

Karlheinz Stockhausen used a maschine (Rotationstisch or rotating table) to rotate sounds around the listeners. A source of noise rotated while microphones positioned around it (four microphones) recorded the sound. Stockhausen was innovative, using the first version of the rotating table to record moving sound for the pieces *Kontakte and afterwards Hymnen,* in cooperation with the West Deutschland Radio (WDR) studios in Germany . A newer version of the rotating table was then developed for *Sirius,* which could turn at much higher speeds than the first version. Stockhausen reports that turning in a circular motion fast enough, the sound becomes machine-like and is not perceived as a performance anymore (Brech & Paland, 2015). Therefore, speed can have drastic implications on sound, not only giving character to the sound regarding its speed or whether it is separated from its single distributed entities, but also defining its apparent size, which is a definite connection of space. When sounds move, they assume a direction or an illusion of it. In a three dimensional space, directions are infinite if the orientation is absent. An orientation can be achieved when fixed references are present (like a sound with a constant position). Dramaturgic and choreographic applications are therefore possible in using combinations to narrate. In Carrè, Stockhausen used these parameters in order to compose his piece. Along with fixed sound sources, he utilized clockwise and anticlockwise movements in his piece (Brech & Paland, 2015, p. 255).

As an example, different waveforms routed to four voltage controlled amplifiers' inputs before going to the outputs is taken. The amplifiers are controlled by envelopes, the attack of which is half a second, and the release or decay duration is one second. Whatever sequence the envelopes are triggered in (rendering rather different trajectories as clockwise and anti-clockwise), and even though the different waveforms differ drastically in sound timbre, movement of the sound source acts as a binding agent, as if all waveforms were a a result of a common event that changed texture according to its position. The completion of the trajectory as a result of choreographed single envelopes contributes to a unity of the different waveforms. As Smalley indicated, spectrum and intensity are intertwined (Emmerson, 1972). This effect is similar to the ones of Bouléz, Xenakis or Brant, who distributed musicians (sources of sound) over a wide area, so that when the musicians played a sentence consequently, the resulting effect is moving sound.

Object Size

In conjunction with the Pitch-Height effect, pitch, size and intensity associations are observed. High pitched sounds for instance, tend to take on a small and light character, which might be related to their elevated position. Low pitched sounds on the other hand favor coming from beneath, a feature also found in heavy objects, which are not easy to elevate. Other associations for lower frequencies are long or elongated (also in duration or the extension of an event), intense, and broadband sounds (Tilley & Cabrera, 2003), as well as slow or slowed down sounds. These associations can be found in real life objects. Large, hollow objects tend to have a deeper sound when excited, while objects that emit higher frequencies tend to be small, compact, and dense. Furthermore, big objects or dwellings need more time for one to maneuver or to move within them, in contrast to smaller architectural or natural structures. This is, however, not a rule of thumb, and surely there are exceptions.

Volume (capacity or the amount of space an object occupies) is dependent on spectral distribution, as well as on intensity. Interestingly and contrarily, when *volume* is measured with varying intensities of sounds, tones with a higher intensity or frequency might appear of greater volume in comparison to those with a lower intensity or frequency. Duration of tones also influences *volume*, since longer tones tend to be perceived as greater in *volume* than shorter tones (Tilley & Cabrera, 2003, p. 2).

Another factor which affects *volume* is the number of locations that contribute to the sound simultaneously. It is more probable for a sound coming out of a larger number of loudspeakers to have a bigger size and volume with increasing width. Therefore, frequency, energy intensity, as well as the number of loudspeakers increase or decrease the size and/or *volume* of a sound object.

Density

The *density* of a sound can be attributed to two different features. First, *density* as in quality (Stumpf's inner moment) of a sound indicates how much of the frequency spectrum the sound occupies and the respective band dynamics that psychoacoustically influence it. A simple comparison would be a broadband sound versus a single sine wave of 800 Hz. The broadband sound (white noise, for example) would appear denser than a single band of 800 Hz. Second, *density* is also represented as a duration and a frequency of occurrence, defined by the rate of the appearance of a sound and the number of spatial locations it occupies. A sound that appears once a minute would be less dense than a sound that appears every second or is sustained the whole minute (*density* in time). Furthermore, a sound that appears in more than a single spatial location is denser than a sound that only occupies a single corner.

Shape

Shape can be represented by a number of things. Dynamics (intensity or sound pressure level) of sound can serve as a good example. The shape of a plucky sound is transient, which might also be described as small, edgy, fast, and vivid. An elongated, sustained tone might be described as long, smooth, slow and seamless. Shape also defines or influences the sound's spatial position. Sustained sounds when sounding alone, pop up in the foreground at first, but eventually blend into the background (even more when other sounds are added) if there is little to no change in its features (in timbre or frequency for example). This is similar to the phenomenon, where a sound that is not altered and maintains a relative constant character tends to be suppressed by the human brain (like an ongoing construction site).

Attention depends on change. Transients might appear exclusively in the foreground, defined by their rapid change of dynamics, demanding attention. When placing transient-like sounds with a sustained sound of little spectral change, the transient-like sounds demand more attention than their sustained counterpart.

Texture and Timbre

In brief, *Texture* is a changeable quality of certain sounds that have far-reaching and profound effects on the sound, as well as on its spatial perception (K. Sharma, 2016). *Texture* is defined by the overall makeup of the spectrum contouring over time with certain fluctuations. Different from *timber*, which is the color of a sound, texture can be a combination of elements that make up a sound or a combination of sounds that make up a *texture*. *Timbre* can be attributed to a single sound (such as a string instrument), which is a salient part of how it sounds, whereby t*exture* can be attributed to multiple sounds (such as how a string section or an orchestra sounds). Both define the structural qualitative composition of sounds.

Changing *texture* or manipulating its aspects over time gives a sound an allure, a term defined by Schaeffer as the characteristic of whether a sound is lively or not. The natural evolution of even the simplest of sounds found in nature undergoes textural transformation, even if it is not lively e.g. wind. In electroacoustic sound, where acousmatic non-naturalistic sounds are produced, changing texture might be analogues to evolving natural sounds or to the gestural action in playing an instrument.

Moving sound objects undergo a spectro-morphological evolution, like fluctuations in the spectral content of moving sound sources due to changes in dynamics, spectral cues, crosstalk cancellation, delay and reflections and so on, caused by movement. Inversely, manipulating spectral content, dynamics, phase, delay and reflections affects movement, as they do in echolocation, which depends on reflections of the surrounding environment (Blesser & Salter, 2009)

Spatial texture is something that Denis Smalley discusses and is related to the *spectral texture*. It is also influenced by frequencies, position, movement, and their superimposition. Frequency, along spectral harmonics, contributes to *spectral texture*. With movement and the use of space, *spatial texture* is enhanced, where for instance, dense flocks of transient sounds yield a different *spatial texture* quality than gradually sustained sounds.

Speed

Rhythm, movement, change, and duration are elements that can be controlled with a modular synthesizer. When the triggering of four envelopes controls the spatial position of a sonic object using voltage controlled amplifiers through amplitude panning (four-channel format), changing the clock speed at which the envelopes are triggered changes the velocity at which this sound object moves (excluding Doppler's effect). This change and generally the speed of a sound has character implications on the sound object. For instance, a fast sound might indicate that the sound object is not heavy, and can move quickly. A slow moving sound might, on the other hand, indicate that the sound object is massive and bulky. Modulating the rate of gates, where at some point the rate becomes very slow and at others fast, slows down or speeds up the movement of a sound, which gives it multiple characters, or a single but malleable character. Furthermore, and similarly controlling the rate of the amount of a signal that goes into spatial synthesizers simulates the rate at which a sonic object enters or exits an enclosure (indicated by illumination of space with sonic energy and its absence). With delays, setting the delay time can determine trajectories of sound objects. This topic has been additionally discussed in the explanation of delay and echo later.

Speed also has implications on *spatial texture, regarding* how fast individual events in a cluster appear and disappear, as well as internally in a sound object, regarding how fast the fluctuations happen and subsequently the resulting *texture* . In changing the spatial texture with speed, the location of a sound object can be modulated in audio rates. This results in the fusion of space and sound together, spreading an object over a larger space, as mentioned previously by (Blauert, 1997). On the other hand, speed of modulation can have major implications on spectral elements. Frequency modulating an oscillator with audio rate levels, as John Chowning did (Chowning & Bristow, 1986) changes the timbral and spectral quality of a sound. Moreover, modulating steady clock rates (clocked clock-modules) might result in rhythmic sequencing and orders of things, which results in a rhythmical spatial texture.

Direction

Direction is a part of sound velocity. It can define the end destination of a sound or be a part of its trajectory. It does not exist without movement or its simulation. When more than one sound is available, a combination of directions can be propagated, which may have a dramaturgical order that, in context, provides details to the sound event. If cardinal directions are not defined and perceivers walk around, a single movement of defined direction would have the same trajectory, but different directions with respect to different perceivers.On the other hand, the absence of direction can be regarded as the diffusion of a sound, where the accuracy of locating the sound movement decreases. This is salient when a dry sound object is moved around on a certain trajectory, and then enters an enclosure, becomes diffused and disappears, while the reverberation tail fades out. This does not mean that reverberation can not have direction; reflections in natural environments are divided into early and late reflections, where early reflections arrive before late ones due to the position of a sound source and the boundaries around it, defining the direction of propagation, and revealing its approximate position within an enclosure. A similar effect can be simulated in modular synthesizers, where one has control over the delay time of reverberators (or pre-delay).

Arena-Spatial Character

Spatial Size

Spatial size can be achieved and controlled through different parameters, such as sound intensity, frequency, timbre/texture, speed, and reflections. What is meant by (arena) spatial size, is how big a space available for utilization is. With frequency and spectral content and pitch-height effect, the median plane spans an additional dimension to the horizontal one. Concerning speed, the faster a sound travels, the more reachable (smaller) a space appears, so that it can oppositely be implied, that the slower a sound travels, the bigger the space appears. Reverberation in a four-channel format would indicate that the perceiver stands in an enclosure, which could define periphery (e.g. spatial and temporal spreading, reflective material, volume and so on). Furthermore, some spatial synthesizers posseses a pre-delay of reverberation feature, which by delaying the early reflections changes the size of enclosure. Furthermore, *spatial texture* can also reveal *spatial size*, in that the larger the number of events (increased density) that emerge, the smaller the space is that fits them (abundance), overlapping in time as well. On the contrary, larger spaces fit a greater number of sounds, in that each sound receives enough space to be perceived, with less overlap and more distance between individual events (increased clarity), thus implying greater space.

Representations of spaces, as indicated before, invoke associations. For example, solitude is associated with hearing large spaces. Intimacy is associated with closeness or tight spaces. Other associations can differ with different perceivers, for example, nearness and smallness can also trigger uncomfortable emotions like claustrophobic feelings and the intrusion upon private space, while large spaces might be associated with loneliness or kenophobia (an abnormal fear of empty spaces). The meaning of these dualities takes form in the appropriate context.

Spatial Feedback

Alvin Lucier's famous art piece "I am sitting in a room" is a good example of *spatial feedback*. Similarly, feeding back a signal in on itself can yield different sounds of different textures. The qualities of the sound are determined by the spatial synthesizer’s parameters, such as which resonances dominate, and at which rate or size, as well as its early reflections delay, and so on. As these parameters can be modulated, changing them constantly results in a spatial body representation with real time alterations. This is almost analogous to constantly and seamlessly changing the characteristics of a concert hall or moving from one to the other.

Arena Spectral Space

Spectral space is the distribution of a single sound's spectrum over and through a particular space. Space becomes then defined by frequencies. Perceivers might indicate that by saying this corner has high frequencies or is sharper and thinner, while the other corner is low pitched, more dense, and round (or other pitch attributes). This is according to K. Sharma (2016), one of 2 divisions of order that he names *the spatialization of the first order*. In brief and as previously described, it is the distribution of frequencies over space, hence the name spectral space. It can also be dependent on time and thus have a sequential nature, increasing dynamic. *The spatialization of the second order* happens when different textures are assigned to different spaces within the frequency range. Here, a step closer towards sculptural sound phenomenon is taken, where space is perceived as a three dimensional setting experienced in sound, depending on the perspectival spatiality and thus, on the listeners´ location and orientation. Spectral space motivates the listener body to space relationship as the audience becomes motivated to experience the space physically, because different areas are occupied by different characteristics. The change of the spectral space structures or its modulation of the spectral space increases complexity, which might invite the audience to experience the spectral space. Fast movements of the allocated frequencies through space might blend some places together and become a greater space, or mask other frequencies that become harder to perceive, inspiring less experimentation.

Dividing space spectrally is an established art researched and practiced by Garriet K. Sharma. Since installation visitors have in common a shared perceptual space (SPS). A relationship of their body to the loudspeaker setup can develop. A stainless-steel tube sculpture made by Eusebio Sempere, mentioned by Blesser & Salter (2009), might be considered as *spectral space*. The structure, depending on the perceivers position, filters a portion of frequencies out, while modifying others. Depending on the perceiver position, the sonic content might be altered.

Modulating the spectral bands that emanate from the loudspeakers results in the change of form of the *spectral space* and makes the spectral structure movable. Without modulation, standing waves might form as a result of the combination of the perceivers position, as well as the surrounding environment. Furthermore, timbral and textural modulations induce a fundamental tonal change which might span over the whole spectral space, uniting it under the modulated parameters. New local spectral space can be formed in that it is detached from the rest of the space using tonal aspects, dynamics, or motion.

Negative space can be defined as the space where the absence of sound is found. In simpler terms, in the four-channel format, spectral space might be achieved by the spectral distribution of a tone over three loudspeakers, while the forth does not emit sound. The spectral space spreads over a certain area, but a void is introduced, where the spectral space is not emitted, similarly to some forms of caves where sound travels into different chambers and comes back deformed at times, or is absorbed by certain material (like plants or moss). Giving places a certain spectral identity reveals the *aural personality* of the space, which makes it easier to identify it.

Delay and Echo

> If there are echoes played very quickly by groups entering each time in different order, and resounding from different points on the stage, the space is "described" each time with a new pattern of these echoes. This is much more interesting than having everybody play together (Boulez 1992: 5). (Harley, 1994, p. 165)

Delay is not only an exceptional tool for spatial sound even in applications with fewer audio channels than four, but also an excellent textural processor. The *Precedence Effect* (Litovsky et al., 1999) in brief is described as the localization of sound, depending on the law of the first wavefront, which states that a sound direction is then determined, when the first sound waves of the first instance of the sound event reaches the ear. The second wavefront determines the spatial tendency of the sound, in that it biases the direction of the sound between the first and second wavefront. Here, time delay is relevant. Delay and echo also have temporal and spatial spreading that was discussed earlier in the thesis. John Chowning claimed that it is more productive, in the spatial audio sense, to have each loudspeaker of the four-channel setup have a separate delay time. This serves to eliminate any stereo signal deficiencies that do not project well on four channels (such as limited spatial panning between two channels, and different simulation of spatial synthesizers. Stereo signals, however, have proven useful for the height phenomenon due to the precedence effect, as well as Blauert's directional bands. When the stereo signal is aligned diagonally between two loudspeakers in the four-channel setup, height effect can be achieved due to the stereo relationship (processing) of the two channels. This works well with reverberation or delay for instance, where both signals are angled in a way that falls in line with Lee (2017)'s experimentation of directional bands and the height effect. Using a switch, the ability to shift the diagonal lines to other lines can be interesting, in contrast to fixed stereo signals.

A quad-delay Max MSP patch was conceived, where a single sound can be delayed four times over four different loudspeakers, independently. A random button was added to randomize the delay times of all four loudspeakers simultaneously. This generates new trajectories produced by motion caused by the delays, everytime the random button is pushed. A second delay reset button was added, which divided a random number by half, two, three, and six corresponding to channels one to four. Furthermore, each of the four channels has a filter (bandpass response) to change the spectral content of the delay, with variable gain, frequency cut-off, and resonance. Midi and audio connectivity are built in for live processing and spatialization, as well as hardware synchronization (clocking). A playback device for premeditated sounds is also available. Pitch shifting, as well as playback speed are independently variable, so that in combination with the filters, it is possible to alienate the output from the original input signal. Such a delay module in modular synthesizers still does not exist, prompting experimentation with such a patch. Having global-delay time unites the space in a four loudspeaker setup, due to their relationship to each other i.e same incremental time change or divisions.

Reverberation

Reverberation and its psychoacoustical ramifications influence human reason and affect. The sonic qualities of a rugged living room elicit the emotional response in many people of being at home or in safety, far from danger (Blesser & Salter, 2009). In contrast, sounding droplets in a hollow cave evoke fright and uncertainty. Just like sounds represent objects, reverberations refer to spaces. In this sense, the space can be taken as the object, and the reverberation is its representation (symbol). Synthesized reverberations are representations of acousmatic spaces, where modifiable parameters indicate the seamless alteration of the space. Acousmatic space hints to sonic spatial representations that are a result of unimaginable enclosures (which one does not see or crossmodally verify). Reflections reveal a space's volume (capacity), wall or boundary material, and shape. It helps in gaining knowledge of the environment. Although a space does not produce sound, it provides necessary information about the sound source, as well as the enclosure itself in the process, e.g. the ratio between direct signal and reflected signal reveals distance between the perceiver and the sound source (Chowning, 2000).

The geometry of the space contributes to its reverberation, while synthesized spaces receive their dimensions from algorithms. Reverberations have attack, sustain, and decay time. Attack time is the duration needed for an enclosure to reflect the sound originating within it (including pre-delay). Decay, the opposite of attack, is the decrease of reverberant energy (Blesser & Salter, 2009, p. 139). Predelay is a controllable parameter in spatial synthesizers, which controls the moment at which the first reflections arrive, as mentioned before. During the sustain period, some frequencies are amplified while others are reduced. The flexibility of modulation in modular synthesizers allows for enclosure manipulation with the change of a knob or control voltage. It is possible, for instance, to modulate the existence of the space itself using the mix (wet/dry) knob. Feedback is as well a modifiable parameter, controlling the amount of trapped energy in space. Tilt filters give the impression that the reflective boundary materials are possible to change, emphasizing high or low frequencies. The size of an enclosure can change, inducing pitch change and timestretch.

The combinations of sound sources and spatial synthesizers can indeed resemble Blesser (2009)'s meta-instrument, in that the space forms a resonant body of the instrument, originating in a whole new instrument. Similarly, space can be used either with or without the sound source, creating new resonances and modifying the sound. This allows sound designers to use space as a sound source itself, and not a conjunction of another sound source. Furthermore, when the dry signal is present and has a trajectory, the reflected wet signal can then be separated from its exciter and receive a different trajectory, forming a seperation of the call-response of a space. Synced symmetrical movements or random trajectories can subsequently be patched to represent a detached sound source and its illuminated space in motion.

In the simulation of virtual sound sources, it can be useful to consider the angle and distance to the listener. As the distance increases between the sound source and the listener, the amount of reflected sound increases and inversely the amount of direct sound decreases. This makes it harder to perceive details of the sound, as it will be diffused due to reflections. Chowning addresses this:

> The amplitude of the direct signal is proportional to 1/distance. Assume the distance from the listener to the point midway between two loudspeakers to be L; we wish to simulate a source at a distance of 2L. The amplitude of the direct signal would be attenuated by ½. (Chowning, 1977, p. 3)

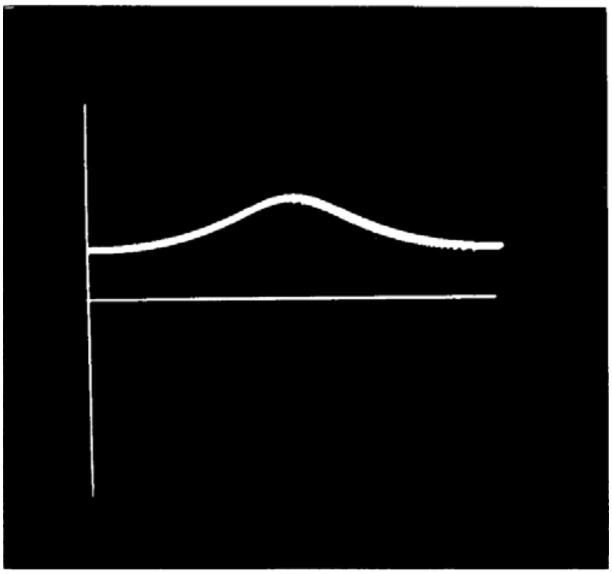

Figure 20. Chowning, J. (1971). Control Function for Amplitude of Direct Signal, 1/D.

The ability to modulate a direct signal to an indirect signal gives direct control over distance within an enclosure, as well as the ability to modulate intensity of the sound object in parallel. Furthermore, changing the material of the walls from stone (reflective) to cloth (amping) changes the spectrum of the reverberation This exceeds the realm of realism and suits an artistic and imaginative approach. When determining angular location using reverberation, it is suggested that every channel has a reverberator with variable and separate delay and gain control, as Ryan Mcgee indicates in (McGee, 2009):

> Reverberation was key in Chowning's implementation of distance techniques. He warns that one must be careful to give each loudspeaker channel its own reverberator with independent delay and gain. Distributing the reverberant signal equally on all channels risks masking the non-reverberant (direct) signal, thus reducing the perception of sound location and direction. (p. 5)

He adds to that:

> In the simplest case some percent of the direct signal is scaled according to 1/√distance and passed to the reverberators equally. Their percent governs the overall reverberation time within the limits determined by the values of the delays and gains of the reverberators themselves.(p. 6)

Supplying all channels with the same ratio of direct signal results in a global reverberation, where all reverberation would be equally distributed, losing the angular temporal and spatial spreading. On the other hand, local reverberation is achieved in addition to preserving angular cues, by adding distance to the equation with respect to the loudspeaker (determined by the angle the object is at), thus making the equation for local reverberation 1 - 1/√distance.

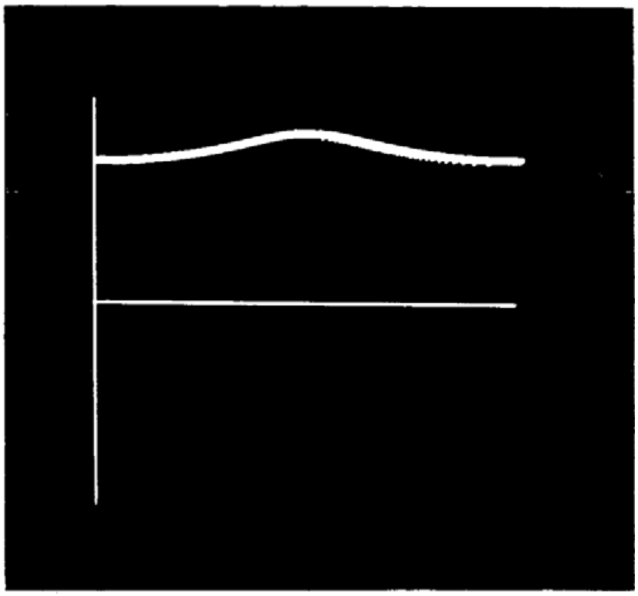

Figure 21. Chowning, J. (1971). Control Function for Amplitude of Reverberant Signal, $1/\sqrt{D}$.

This indicates that with increasing distance from the sound source (narrower angle), the reverberation becomes more localized (Chowning, 1977, p. 3). With this knowledge, global and local reverberation can be utilized contextually and artistically. Determining how much of the dry signal reaches the reverberators can be set manually or by control voltage modulation in a synthesizer. Moreover, switching sequentially or randomly between global and local reverberation can be intriguing, as the possibility of both options exists.

Convolution Reverb is a digital application enabling the import of a real space impulse response (in short "IR", recorded using a sweep or an impulse) which is then processed by a convolver, recreating the space in which this impulse response was recorded. A stereo impulse response recorded in a space has differences in both channels (like in ITDs and ILDs). This renders these convolved signals realistic sounding, as it is a mimic of binaural hearing. In the application of a four-channel format, using convolution reverb enhances the spatial impression of reverberation. Loading sounds instead of impulse responses into the convolver renders it as a sound design resonator. This application of convolution enables the presentation of e.g. a dog bark as space.

Furthermore, reverberation in an installation depends on the space in which the installation takes place. Together, the physical and the virtual fuse into one another, invoking a third space.

Parameter Control (Modulation Character)

Modulating the Modulator

Frequency, pulse width, and polarity are but a few parameters which when modulated, affect the end modulation result. For instance, a low frequency oscillator (LFO) which has a frequency modulation (FM) input would increase or decrease its oscillation speed depending on the signal at the FM input. This results in complex, gestural modulation and waveshaping possibilities, which in turn can modulate timbral and spatial characters. Similarly, a trigger module that is clocked by an external clock results in rhythms determined by the wave-shape of the rate-modulation input signal. In general, these modulations induce a life-like feature or an allure to other static sources, where the factor of change is also changed itself. In nature, it is difficult to find a steady state of things such as a pure sine tone oscillation or a constant rate at which events happen.

Speed is always in combination with time, for example in physics, speed is the distance covered in a certain amount of time, which is a relational binding of physical and temporal dimensions. This is also the case with the speed of moving sound objects, which theoretically can be measured as well (distance over time). Since the triggering of envelopes that control a signal amplification over four loudspeakers in a four-channel format simulates movement through amplitude panning, the speed at which envelopes are triggered determines the speed of the sound object travels. Modulating the rate of the device which triggers the envelopes results in modulating the speed of the movement of the object, so that when the rate increases, speed of motion increases. Similarly, modulating the rate of repetitive transient sounds without necessarily having traveling motion results in ample modulation of the events, so that when the rate of the clock is high, the appearance of the sound events increases, and vice versa.

Moving a sound using only voltage controlled amplifiers and the help of a modulator is an excellent way to make use of spatiality in a modular synthesizer. Substituting velocity of a moving sound object can also be achieved using this concept. In this case, Doppler's effect can be achieved when an attenuated version of the control voltage that is utilized for motion is also used to modulate the pitch. Chowning (1971) mentions: "The simulation of the Doppler effect is achieved simply by computing the distance D from the subject to the apparent source and making change in frequency proportional to dD/dt".

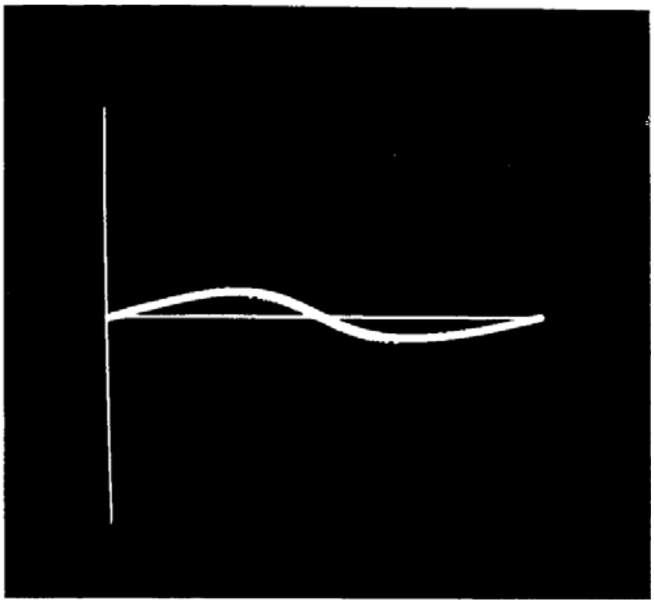

Figure 22. Chowning, J. (1971). Control Function for Doppler Shift, dD/dt.

Patch Applications

Generative Sound

To explore the complexity and unpredictability that can be achieved with modular synthesis, we take a closer look at generative sound, which is the method used in the framework of the installation. Probabilistic models, such as the Markov's chain, named after Andrey Markov, comprise realistic probabilities. These models simulate a real probabilistic sequencing in which events occur, based on the present and past states. Iannis Xenakis has also worked with similar models, namely the Stochastic. Alan Turing,who helped to decipher World War II codes, also developed an abstract computing method which alters symbols on a tape band according to a set of rules. Although it is a rather elementary and straightforward calculation, it is capable of handling complex algorithms. Here, it is important to create a probabilistic or randomization model in a *controlled* manner. John Cage is another artist who dealt with randomisation in music, as the American composer also based his music on the book I Ching, combining and altering sequences.

In definition, generative sound is sound which is generated according to controlled random variables. This can be for instance a non-repetitive melody in a single musical scale. Modules that can help in the creation of generative sound are turing machines, sample and hold or track and hold modules, voltage samplers, Logic, quantizers, clocks, and LFOs.

For example, resetting a free oscillating LFO with random triggers yields steadily oscillating outputs that are restarted randomly. Another example is using a sample and hold voltage (fed with random voltage) to frequency modulate LFO frequency. The square wave output of the LFO, can be routed to the trigger input of the sample and hold module. This way, both modules influence each other, in that when an oscillation of LFO cycle is complete, the sample and hold is triggered by the LFO square wave, generating a new random voltage, which in turn changes the oscillation speed of the LFO, which again triggers the sample and hold module, changing the LFO's frequency again. At every new voltage, a new oscillation speed is generated, thus changing the trigger time of the sample and hold. The remaining outputs of the LFO can then be used in the patch to modulate all sorts of parameters, ranging from spatialisation to texture, even both together. This is a classical example of a mutual and self influencing modulation programming.

Complex or Multi-Level Panning

An intriguing way to make complex spatial modulation is to modulate amplitude panning modulation. Imagine four envelopes responsible for the modulation of the sound pressure levels of four signals over four-channels. When the envelopes are triggered and have somewhat of an overlap, the result is a continuous trajectory of the sound. Attenuating these envelopes with control voltage using a seperate set of four VCAs changes the scale or amount of envelope amplification. This results in the increase and decrease of amplification of the envelopes, which in turn affect the sound pressure level of the moving sound, and consequently its position and velocity. This two-layer panning method is a combination that results in a complex panning. More layers can be subsequently added, increasing modulation complexity.

Spectral Space

Dividing space is achievable with four filters or resonators, where each of the four filter outputs is routed to a seperate loudspeaker. Surely, modulations would make the spectral space more dynamic, in that the frequencies radiated would change depending on the modulation, which is equivalent to moving the spectral space with respect to the listener, who is also free to move around with respect to the loudspeakers.

Multi-level Spatial Composition

Having multiple layers of sounds (Flocks or clusters) of sound increases complexity. Sounds and/or flocks of sound can be differentiated by different parameters with which an identity is given. A good example is sound duration that is determined by an amplitude envelope. Consider envelope durations divided into three categories: short, medium, and long. Sounds that have similar durations would be grouped automatically by the perception as belonging to the same category (short sounds), due to the similarities in the nature of their amplitude contour. Their perception might be nearer and more attention demanding than ones that swarm and float in the background. Other identities can be given with frequency range, timbre or texture (spectral content), or density (rate of triggered event in a cluster) for example.

Yet, another form of clustering can be achieved through spatiality or spatial character, even if the sounds have separate spectral and dynamic content. Speed of movement is a category of such, where sounds of similar velocities are perceived to be related and thus are grouped together as a flock.

Spatial separation can also be achieved, where the dry to reflected signal ratio of reverberators or delays is differently set for some sounds in comparison to other sounds or their relative clusters. Amplitude intensity which resembles nearness at high intensities and farness at lower ones can contribute to multi-level spaciousness.

Having these different categories of identification enables the sound designer to decide how sounds and sound clusters behave and in which sequence they appear. Giving sounds an identity characterized by the ones mentioned above, enables the simultaneity of sounds that are separable because of their qualities, including spatial character. Spatial character is much more limited with fewer channels, and is at most confined with a single mono channel.

Achieving this multi-layered composition with a single sound is possible when a quad-delay is available. For instance, using a four-channel VCA routed respectively to the loudspeakers, with a sum output which is routed to the input of the quad delay. The quad delay is then routed to four filters, which in turn, are routed to the outputs again. Now, as one of the four VCAs amplifies its signal, the same signal is sent to the quad-delay as a result of routing the sum output of all VCAs to the delay module. Here, the multi-layer separation happens in that the dry signal receives a specific place (like a corner or where the corresponding loudspeaker is), while the delayed and filtered signal follows and travels in a trajectory depending on the delay times (where it might be time stretched or receive a different texture). Additionally, one delay output can then be sent to four bandpass filters, which then divide the single delay output into spectral space. Furthermore, global or local reverberation can be applied to the bandpass filters (and for instance, having its parameters like size, feedback and wet/dry mix modulated). This creates a multi-level spatial composition. Their alteration and appearance in time can then be dramaturgically or randomly sequenced.

Spatial ADSR

A spatial ADSR envelope outlines the stages of spatial distribution of an event. It defines a sequential order that determines spatial happenings. This can be imagined as follows: Take a transient-like sound which over the four-channel setup appears in a moderate density (can be heard simultaneously at different locations). This can be any sound from note to noise. The first stage can be defined as the attack stage of the spatial order and the offset of a spatial sonic event. Using amplitude panning, the dry, unmodulated sound is moved around. The second stage would be decay, which changes the spatial distribution of the sounds, and other aspects like density as well as sonic qualities like pitch or timbre, to induce spectral alteration. This can be for instance, the change of ratio between dry to wet signal over global or local reverberation, making it diffused and reflected, paralleled by the timbral and spectral change. The shape of the dynamics envelope can change as well from linear to logarithmic or exponential, or change duration from short to longer. After decaying, the sustain stage, which is the third stage emerges. Here, yet another change in spatial and textural qualities happens: transforming the diffused signal into spectral space, where from the broad reverberation signal only some spectral bands materialize to fill the space variously. Lastly, and once more the spatial envelope steps into the last stage, the release stage. Here, this involves fading the events into disappearance, while changing its sonic and spatial qualities anew. From spectral space, and using the amplitude technique, the spectral bands can be brought to a pulsating event to mimic the transient-like contour that is achieved in the first phase, and routing the sum of all signals to a delay module. This happens while changing the frequency of the sound, as well as changing the dry to wet ratio of global delay taplines.

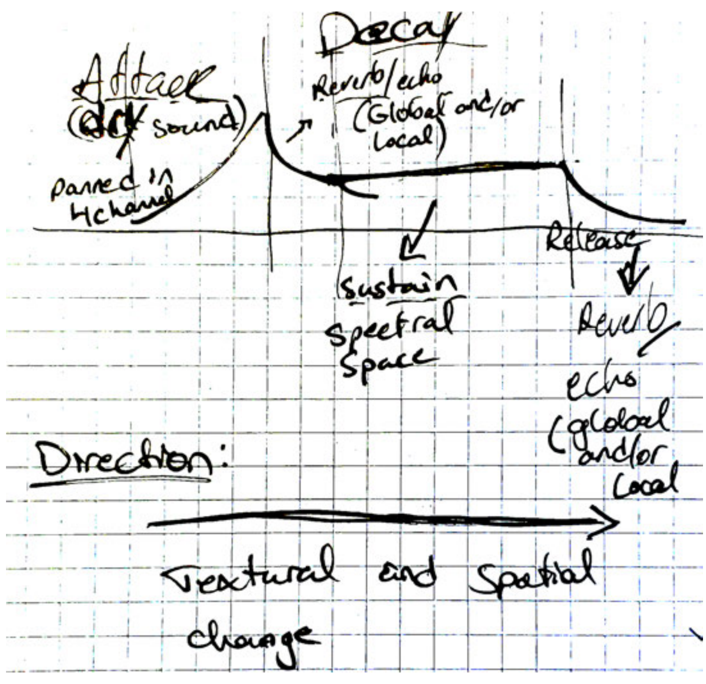

Figure 14. A Sketch of the spatial ADSR Model.

The spatial envelope in this case dictates the allure of sonic happenings. Some modules provide single stage outputs (a gate at the completion of every stage of the ADSR), which fits the ADSR spatialization model perfectly. The ADSR envelope can take on various speeds through time, ranging from milliseconds to hours.

Conclusion

Immersion and Envelopment

One of the strongest envelopment effects of sound is reverberation. As Blesser (2009) reported:

> with reverberation, a sound wave is equally likely to arrive from the left, right, front, back, above, below. When a sound field is fully diffused, uniform, and random, you cannot detect the location of its source because your auditory experience is identical in all directions. (p. 145)

Achieving immersive envelopment has immense dependence on signal correlation degree. Decorrelated signals of any kind, including reverb signals, have a position easier to locate, and lend a realistic feel. Brech & Paland (2015) quote one Columbia engineer working on fine tuning the acoustics of the Greek Orthodox Church on East 30th Street in New York in the 1940s and turning it into a recording studio: "A string of very distinct sixteenth notes would come back as a smear". The engineer proceeds that it felt as if one "were *immersed* in reverberation". The term *immersed* was used to describe the situation in which reverberation of the venue was created.

Envelopment is enhanced by movement, which in turn enhances immersion. The relationship between immersion and envelopment is one of active influence, where one impacts the other. For instance, the speed at which a sound moves results in perceiving a sound either as a single element in space moving slowly, and is immersive and easy to localize, or at high speeds, where diffusion occurs and envelopment is increased.

Rumsey in a 2002 article suggests, on the other hand, that immersion has two attributes, presence and envelopment. Presence is meant in the sense that one finds him or herself in a room as an example, while envelopment is the effect of having more than one sound source distributed around the perceiver (K. Sharma, 2016, p. 78). It is also reported that immersive envelopment, which is also called listener envelopment, depends on reflections of sound. There has been a distinction between "room envelopment" and "source envelopment" as well, which respectively refer to reflections and sound source surrounding.

Connecting Spaces and Travel

Having a number of spaces in a single installation is metaphorically equivalent to changing perspectives or spaces physically. This associativeness can be used to build up or break down spatial structures and metaphors. Changing the parameters of reverberations and other spatial synthesizers using modulation induces alterations to enclosures or resembles the malleability of space, changing boundary material, size, shape, energy reflection and so on. In a synthesizer environment, this can be achieved seamlessly, in time or out of it, randomly or purposefully. The change in space might indicate the change in perspective, as would a sound moving around the listener. The following questions arise: where am I with respect to the sound source? Is the sound source in front of me or am I behind the sound source? These questions can help in the creativity of sonic situations and metaphors in order to communicate ideas.

Animation, Allure, and Emotion

Using space in sound adds another channel of information, conveying and portraying cues that affect the cognition and emotion. Depending on context, for instance, it can be assumed that “up” is high and superior while “down” is low and inferior, in reference to spatial position. High can also be inferior and small while low can be big and morbid. This is almost always dependent on the context in which these associations take place, as mentioned before.. These channels can hold information about a possible cause of the emitted sound and how it was, if it was, processed. So is, for instance, a musical phrase or gesture, consciously or subconsciously, imagined being performed by a player physically caused by the movement of body parts and the instrument as an extension to it. This is also called *First Order of Surrogacy* according to (Emmerson, 1986, p. 82). As a familiar sound (like a musical instrument) is deformed or manipulated, yet still maintains a recognisable part of its identity, this would be the *Second Order of Surrogacy*. In both cases the allure of the sound exists, as the cause of the sound is a gesture resulting from the playing of an instrument. This information is embedded in the sound within many characteristics, like timbre, dynamics, movement and so on. Beyond the second order comes an abstract form of sound, of which the cause is hardly recognisable, hence the name *remote surrogacy*. Physical causation of the sound, or the sound producing object. in this case. is perceptually altered and becomes based on pure psychological explanation (Emmerson, 1986, p. 83). Although its cause is not clear, allure can still be maintained, in which the same channels of gesture-carrying information are retained. Timbre, dynamics, and location for example can still be communicated and indeed point at some kind of causation of why they sound the way they do, even if ambiguous. Moreover, there is a fourth type of surrogacy mentioned by Emmerson (1986), namely the *dislocated surrogacy*. At this point, sounds are fully dissociated from their source context, which is described by Emmerson (1986) as follows: "Our remarks concerning spectro-morphological design, the implications of motion pacing, and a multi-level focus are all symptoms of unviable, *dislocated surrogacy"*. A question one might be led to ask when listening to acousmatic sound is whether the sound could be tangible or not. Motion does indeed add a live-like accentuation. Structures, furthermore, can be separated into two different categories: *gesture-carried* or *texture-carried,* where in the former movement dominates attention, while *texture-carried* is dominated by the constitutional change of inward texture, like frequency change. This is similar to Stumpf's *inner* and *outer-moments*. The two categories share attention, only this happens separately. When the gesture is slow enough, attention shifts to textural changes. When texture is passive, and movement is at hand and attention shifts to motion.

When listening to two or more sounds in a mono, it can be intimidating to differentiate between them when they are played simultaneously and if they occupy a similar spectrum domain. Spatially separated sounds though, are much easier to perceive separately. Hence, it can be assumed that with spatiality, sound clarity is enhanced. Stockhausen used spatial separation in order to achieve coexistence between two layers of sounds to be perceived as individual assets (Brech & Paland, 2015, p. 179). Brech & Paland (2015) also writes about the piece Diario Polacco from Luigi Nono in 1958: "Luigi Nono also used an arrangement of the positions of the musicians in the exterior space as a listening aid to facilitate the perception of characteristic sounding quality". The use of space also allows the elongation of sounds as they might not disturb the other ongoing sound objects or events, when they occupy a different space, and therefore have an effect on time.

The composition in a four-channel format in a modular environment prompts thinking of space during conception alongside the sound. Before spatializing a sound, some questions might arise. How does the aspired sound behave spatially, and why would this sound of certain frequency or pitch, texture, and dynamics behave the way it behaves? What is the desired goal and in which level does it reside in the multilevel composition? Does it have complex trajectories or does it travel over a line between two points? There is no right or wrong in which process comes first (spatialization or sound design), only which approach best serves the end result. Certainly, one can start with spatialisation of, for example, a sine wave, in that it receives a velocity over a certain trajectory, and using techniques like additive synthesis, develop the timbre and texture that caters to its spatial behavior. It is for example already useful for sound objects that travel in the vertical dimension, contain or pass through the range of frequencies that are responsible for the elevation of sounds in psycho-acoustics. Moreover, how does the mutual influence of spatial character and sonic identity take form? The spatial dimension and allure are enhanced when the relationship of space and spectral change have a commonality, like a single modulating source that modulates position, amplitude and spectral bandwidth, for example.

Cultivating the relationship between the inner modulations and fluctuations and their outer spatial contour is productive. Sound wins allure and a life-like quality, when its inner features are projected outwards in spatial character.

Furthermore, sounds or flocks of sounds, are sounds of similar qualities that can be grouped together. Relationships arise between entities, as closeness to one another might portray intimacy and fusion, while farness might evoke alienation. A call-response ordered in time and by spatial structure, differentiated by timbral difference, and spatial division can develop leading up to dramaturgical sequences. Having one static sound while the other moves around it achieves a concept of a static reference to a moving object. This proves to be helpful in the building of spatial structures, and serves to anchor the attention to a recurring event, which establishes an assurance that it will reoccur, while other things might be changing with uncertainty.

Outcome

Using a modular synthesizer in a four-channel format as a tool to produce spatial sound content can be enhancing to the medium. This is due to the mutual influence mechanisms that take place when composing sound and space together. Immersion and envelopment increase drastically and sounds acquire an allure, a life-like or even human-like character which is formed and shaped within and because of space. Space also becomes an instrument of itself, after it might have been a meta-instrument together with the sound source illuminating it. It can induce feelings and thoughts in forms of triggered associations. Religious spaces like churches, mosques, and temples tend to be big hollow structures that result in long reverberation, which motivates a sense of awe at the perceiver level. This can also lead to a sense of loneliness due to the vastness of the space, implying the smallness of the perceiver. The projection of inner processes of sound qualities, like timbral and textural fluctuations, outwards as motion can prove rewarding in vitalising sound and enhances the allure. Replicating, changing, and alienating spaces is now possible due to the technological advancements, where "architecture" can be malleable. Space can be achieved with sound in several methods, like spatial panning, reverberations and delays, spectral content, and spectral spaces. Auditory spatial awareness is the internal experience of the external environment, and similarly, the experience of the external spatialisation of internal sonic phenomena. A creative installation with generative synthesis is realized.

Bibliographie

Ableton, & Dieter Döpfer. (2015, February 18). *Interview with modular synthesis guru Dieter Doepfer*. Ableton; Ableton Homepage. https://www.ableton.com/en/blog/dieter-doepfer-completing-the-circuit/

Adajian, Thomas, & Edward N. Zalta. (2018). *The Definition of Art*. Stanford Encyclopedia of Philosophy; Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/art-definition/

Adelbert W. Bronkhorst. (2000). *The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions* (Issue Acta Acustica united with Acustica 86(1):117-128). https://www.researchgate.net/publication/230739432_The_Cocktail_Party_Phenomenon_A_Review_of_Research_on_Speech_Intelligibility_in_Multiple-Talker_Conditions

Bader, R. (2013). Sound - Perception - Performance. *Current Research in Systematic Musicology*. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-00107-4

Badura, J., Dubach, S., Haarmann, A., Mersch, D., Schenker, C., & Pérez, G. T. (2015). *Künstlerische Forschung. Ein Handbuch*. https://books.google.com/books/about/K%C3%BCnstlerische_Forschung_Ein_Handbuch.html?hl=&id=yobGrQEACAAJ

Batchelor, P. (2015). Acousmatic Approaches to the Construction of Image and Space in Sound Art. *Organised Sound*, *20*(2), 148–159. http://dx.doi.org/10.1017/s1355771815000035

Bhandari, Prabin. (2017). Interactive Voice Response with Speech Recognition. 10.13140/RG.2.2.16148.60802.

Blauert, J. (1969). *Sound Localization in the Median Plane* . Acta Acustica united with Acustica. 22. . https://www.researchgate.net/publication/263716618

Blauert, J. (1997). *Spatial Hearing*. MIT Press. https://books.google.com/books/about/Spatial_Hearing.html?hl=&id=wBiEKPhw7r0C

Blesser, B., & Salter, L.-R. (2009). *Spaces Speak, are You Listening?* Mit Press. https://books.google.com/books/about/Spaces_Speak_are_You_Listening.html?hl=&id=YusCEAAAQBAJ

Brech, M., & Paland, R. (2015). *Kompositionen für hörbaren Raum / Compositions for Audible Space*. transcript Verlag. https://play.google.com/store/books/details?id=xDiyCwAAQBAJ

Bryant, D. (1990). *The 'cori spezzati' of St Mark's: myth and reality*.

Buchanan, F., & DeVito, J. A. (2016). *Communication Skills in social work and human services*. Pearson Australia for the University of South Australia.

Capanna, A. (2015). *Conoids and Hyperbolic Paraboloids in Le Corbusier's Philips Pavilion*. https://doi.org/10.1007/978-3-319-00143-2_25

Chowning, J. M. (1977). The Simulation of Moving Sound Sources. *Computer Music Journal*, *1*(3), 48. http://dx.doi.org/10.2307/3679609

Chowning, J. M. (2000). Digital Sound Synthesis, Acoustics, and Perception: A Rich Intersection. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*. http://www.music.mcgill.ca/marlonschumacher/wp-content/uploads/MUCO541/Readings/Chowning_DAFX2000.pdf

Chowning, J. M., & Bristow, D. (1986). *FM Theory & Applications*. Hal Leonard Corporation. https://books.google.com/books/about/FM_Theory_Applications.html?hl=&id=I48YAQAAIAAJ

Dalton, P., & Fraenkel, N. (2012). *Gorillas we have missed: Sustained inattentional deafness for dynamic events*. Department of Psychology, Royal Holloway, University of London. https://doi.org/10.1016/j.cognition.2012.05.012

de Boer, K. (1947). *A Remarkable Phenomenon with Stereophonic Sound Reproduction*. Philips Tech. Rev.

Dickreiter, M. (1982). *Handbuch der Tonstudiotechnik*. Walter de Gruyter. https://books.google.com/books/about/Handbuch_der_Tonstudiotechnik.html?hl=&id=HyciAAAACAAJ

E. Colin, C. (1953). *Some Experiments on the Recognition of Speech, with one and with two Ears*. The Journal of the Acoustical Society of America.

Emmerson, S. (Ed.). (1986). *The Language of Electroacoustic Music*. Palgrave Macmillan UK. http://dx.doi.org/10.1007/978-1-349-18492-7

Emmerson, S. (1998). *Aural landscape: musical space*. Organised Sound, 3(2), 135-140. https://doi.org/https://doi.org/10.1017/S1355771898002064

German Keys Magazine, edition 12/1997, starting at page 152. (1997). *Doepfer-Story - Der Modul-Mogul*. Doepfer Musikelektronik GmbH; German Keys Magazine. https://doepfer.de/hist_e.htm

Görne, T. (2017). *Sounddesign*. Carl Hanser Verlag GmbH Co KG. https://play.google.com/store/books/details?id=rOUlDwAAQBAJ

Hannula, M., Suoranta, J., & Vadén, T. (2014). *Artistic Research Methodology*. Peter Lang US. http://dx.doi.org/10.3726/978-1-4539-1308-6

Harley, Maria Anna. (1994). Space and Spatialization in Contemporary Music. In *Space and Spatialization in Contemporary Music - History and Analysis, Ideas and Implementations*. McGill University, Montreal.

Holmes, T. (2020). *Electronic and Experimental Music*. Routledge. http://dx.doi.org/10.4324/9780429425585

Itoh, M., Iida, K., & Morimoto, M. (2007). Individual differences in directional bands in median plane localization. *Applied Acoustics*, *68*(8), 909–915. http://dx.doi.org/10.1016/j.apacoust.2006.08.001

James, W. (1950). *The Principles of Psychology* (Henry Holt and Company, 1890, Vol. 1). Dover Publications, Inc. https://books.google.com/books/about/The_Principles_of_Psychology.html?hl=&id=jRitxrJX4fIC

K. Sharma, G. (2016). *Composing with Sculptural Sound Phenomena in Computer Music*. University of Music and Performing Arts Graz.

Lattal, K. M. (2013). Pavlovian conditioning. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, & K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 1. Methods and principles* (pp. 283–306). American Psychological Association. https://doi.org/10.1037/13937-013

Lee, H. (2016). *Perceptual Band Allocation (PBA) for the Rendering of Vertical Image Spread with a Vertical 2D Loudspeaker Array*. Applied Psychoacoustics Laboratory (APL), University of Huddersfield, Huddersfield. https://doi.org/https://doi.org/10.17743/jaes.2016.0052

Lee, H. (2017). Sound Source and Loudspeaker Base Angle Dependency of Phantom Image Elevation Effect. *Journal of the Audio Engineering Society*, *65*(9), 733–748. http://dx.doi.org/10.17743/jaes.2017.0028

Litovsky, Ruth & Colburn, H. Steven & Yost, William & Guzman, Sandra. (1999). The precedence effect. *The Journal of the Acoustical Society of America*. 106. 1633-54. 10.1121/1.427914.

McGee, R. (2009). *John Chowning: Overview, Techniques, and Compositions*. The University of California, Santa Barbara.

Metz-Göckel, H. (2016). *Gestalttheorie und kognitive Psychologie*. Springer Fachmedien Wiesbaden. http://dx.doi.org/10.1007/978-3-658-12666-7

Moretti, L. (2004). *Architectural Spaces for Music: Jacopo Sansovino and Adrian Willaert at St. Mark's*.

Oliveira, N. D., Oxley, N., Petry, M., & Archer, M. (1994). *Installation Art*. Smithsonian Inst Press. https://books.google.com/books/about/Installation_Art.html?hl=&id=QcTpAAAAMAAJ

Paterson, J., & Lee, H. (2021). *3D Audio*. Routledge. https://play.google.com/store/books/details?id=1WQ0EAAAQBAJ

Plessner, H. (1982). *Mit anderen Augen*. https://books.google.com/books/about/Mit_anderen_Augen.html?hl=&id=ky_lAAAAMAAJ

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*(3), 278–285. http://dx.doi.org/10.1037/h0072651

Pratt, Carroll C. (1929). *Spatial Character Of High And Low Tones*. Harvard University.

Rosenthal, M. L., & Rosenthal, M. (2003). *Understanding Installation Art*. Prestel Pub. https://books.google.com/books/about/Understanding_Installation_Art.html?hl=&id=ApVQAAAAMAAJ

Smalley, D. (1997). *Spectromorphology: explaining sound-shapes*. Department of Music, City University. https://www.semanticscholar.org/paper/Spectromorphology%3A-explaining-sound-shapes-Smalley/6fe2275e2f543d239cd4587b6b577e4f9bcfd357

Smalley, D. (2007). Space-form and the acousmatic image. *Organised Sound*, *12*(1), 35–58. http://dx.doi.org/10.1017/s1355771807001665

Stockhausen, Karlheinz. (1972a). *Four Criteria of Electronic Music (Kontakte)* (Vol. 1). YouTube. https://www.youtube.com/watch?v=7xyGtI7KKIY

Stockhausen, Karlheinz. (1972b). *Four Criteria of Electronic Music (Kontakte)* (Vol. 2). YouTube. https://www.youtube.com/watch?v=ZzDhg-MoHlo

Stockhausen, Karlheinz. (1981). *Tuning In- Karlheinz Stockhausen (Maconie, 1981)*. YouTube. https://www.youtube.com/watch?v=qGnkZnm9MPw

Stumpf, C. (1883). *Tonpsychologie*. S. Hirzel. https://books.google.com/books/about/Tonpsychologie_von_Dr_Carl_Stumpf.html?hl=&id=7nxftwAACAAJ

Tilley, S., & Cabrera, D. (2003). *Vertical Localization and Image Size Effects in Loudspeaker Reproduction*. AES 24th International Conference on Multichannel Audio. https://www.aes.org/e-lib/browse.cfm?elib=12269

Tversky, B. (2003). Structures Of Mental Spaces. How People Think About Space. *Environment and Behavior 35(1)*, 66–80. https://doi.org/10.1177/0013916502238865

Valéry, P. (1937). Rede über die Ästhetik. In *Trivium - Fondation Maison des Sciences de l'Homme, 2010, 6(2010)*. Revue franco-allemande de sciences humaines etsociales - Deutsch-französische Zeitschrift für Geistesund Sozialwissenschaften. http://trivium.revues.org/

Valle, A., & Lombardo, V. (2022). *The Virtual Electronic Poem project*.

Waller, D., & Nadel, L. (Eds.). (2013). *Handbook of spatial cognition.* American Psychological Association. http://dx.doi.org/10.1037/13936-000

Ich versichere, die vorliegende Arbeit selbstständig ohne fremde Hilfe verfasst und keine anderen Quellen und Hilfsmittel als die angegebenen benutzt zu haben. Die aus anderen Werken wörtlich entnommenen Stellen oder dem Sinn nach entlehnten Passagen sind durch Quellenangaben eindeutig kenntlich gemacht.

Hamburg, den 14.07.2022 [redacted]

Ort, Datum