Kurs: **EA2003 Självständigt arbete i musikproduktion, 30,0 hp**

2022

Konstnärlig masterexamen i musik 120,0 hp

Institutionen för Musik- och medieproduktion

Handledare: Hans Gardemar, Jan-Olof Gullö

---

Henry Mikkonen

**Lost in Space**

Three Case Studies in Music Production Using Immersive Audio

Skriftlig reflektion inom självständigt arbete

Abstrakt

Detta examensarbete innefattar en utforskning om mitt arbete med immersivt ljud och hur det har påverkat min konstnärliga praktik som musikproducent. Utforskningen görs genom en analys av tre delstudier bestående av en 360-video som spelades in i S:t Jakobskyrkan i Stockholm, en VR upplevelse i en virtuell kopia av Nathan Milsteinsalen på Kungliga Musikhögskolan, samt en komposition som producerades med Dolby Atmos. Slutsatsen som dras är att arbetet har påverkat min konstnärliga praktik genom att ge mig nya insikter på komponerande, inspelning, och mixning. Slutligen ges förslag på fortsatta studier om audiovisuella metoder för immersiv musikproduktion samt nya infallsvinklar på stereomixning som är informerade av immersivt ljud.

Nyckelord: Musikproduktion, immersivt ljud, Ambisonics, Dolby Atmos, Virtual Reality (VR), 360-video, psychoakustik.

Abstract

This thesis examines how working with immersive audio has affected my artistic practice as a music producer. It does so by examining the insights gained during the production of three case studies consisting of a 360-video recorded at Saint James’s Church in Stockholm, Sweden, a Virtual Reality (VR) experience in a 3D replica of Nathan Milstein Hall at the Royal College of Music in Stockholm, as well as a composition produced using Dolby Atmos. The paper concludes that working with the case studies has changed my artistic practice by leading to new insights in the areas of composing, recording, and mixing. The paper also suggests future studies into the use of audiovisual approaches to immersive music production as well as considerations for stereo mix techniques based on insights gained from immersive audio.

Keywords: Music production, immersive audio, Ambisonics, Dolby Atmos, Virtual Reality (VR), 360-video, psychoacoustics.
# Table of Contents

**Introduction** ....................................................................................................................................... 1

Aim and Purpose ..................................................................................................................................... 1
Research Question .................................................................................................................................. 1
Constraints .............................................................................................................................................. 2

**Previous Studies in Immersive Audio Related to Music Production** ................................. 2

Technical Research .................................................................................................................................. 2
Historical Approaches to Sound in Acoustic Spaces ................................................................................ 4
Immersive Music Production at the Intersection of Acoustic Practice and Virtual Spaces .......... 5
Ethical Considerations ............................................................................................................................. 5

**Three Case Studies in Music Production using Immersive Audio** ................................. 6

Case Study No. 1: 360-video Concert at Saint James’s Church ............................................................... 6
   *Method and Results* ............................................................................................................................ 7
   *Analysis* ............................................................................................................................................. 13
Case Study No. 2: VR Concert in Nathan Milstein Hall .......................................................................... 14
   *Method and Results* .......................................................................................................................... 15
   *Analysis* ............................................................................................................................................. 16
Case Study No. 3: Dolby Atmos Music Project ...................................................................................... 17
   *Method and Results* .......................................................................................................................... 18
   *Analysis* ............................................................................................................................................. 19

**Final Reflection** .............................................................................................................................. 21

Future Studies ....................................................................................................................................... 22
   *Audiovisual Immersive Production* .................................................................................................. 22
   *An Immersive Approach to Stereo Mixing* ..................................................................................... 23

**References** ........................................................................................................................................ 28
Introduction

The term immersive audio refers to audio technologies that enable the playback of sounds in three dimensions, as opposed to the horizontal playback of traditional stereo and surround sound setups. Two popular immersive audio technologies today are Dolby Atmos and ambisonics. Dolby Atmos emerged in the field of film production and is now also being used for music production, while ambisonics emerged as a surround format in the 1970s and is now commonly used in sound installations, 360-videos, and VR productions.

Aim and Purpose

I have created three case studies consisting of a 360-video and a VR concert experience utilizing ambisonics, as well as a composition using Dolby Atmos. The purpose of these case studies has been to explore how working with immersive audio has impacted my artistic practice as a music producer.

My aim has been to contribute with knowledge regarding the creative aspects of working with immersive audio, as much of the currently available information is of a technical nature.

A secondary aim has been to contribute with a perspective on music production where immersivity has been an expectation from the outset. In other words, the idea that the music will be presented using immersive audio has been a given, and has influenced the entire creative process, from composition to recording and mixing. I am highlighting this perspective to counterbalance what I see as tendencies to focus on upmixing stereo material and treating immersive audio as an extra deliverable on the side of a stereo release. In my view, this kind of approach limits the potential uses of immersive audio and may in the long run prevent its widespread adoption due to a lack of compelling use cases for the technology.

Research Question

- What new insights have I gained into the composing, recording, and mixing processes through working with immersive audio?
**Constraints**

The scope of this thesis is to describe how working with immersive audio has affected my artistic practice as a music producer, thereby contributing to an emerging field which seeks to explore artistic uses for immersive audio in music production. I have not intended to endorse any one kind of immersive audio technology over another nor provide explanations of the differences between them. Instead, the focus of my work has been to encourage ways to think immersively when engaged with the creative and practical work of music production.

While my work has implications for an audiovisual approach to the music production process as well as a shift to an immersive audio perspective in stereo reproduction, these aspects are not fully explored in this paper but would be well served by future study.

**Previous Studies in Immersive Audio Related to Music Production**

In surveying the existing research in the field of immersive audio, I have primarily come across technical research which include studies into the psychoacoustic aspects of immersive audio, explanations of the technology behind various immersive audio technologies and suggestions for practical application.

To better understand creative uses for immersive audio, I have also studied historical approaches to the use of acoustic space by composers.

To connect the technical aspects of immersive audio with the creative possibilities highlighted by historical approaches to acoustic space, I drew from Blesser and Salter’s (2006) concept of aural architecture, to suggest that music production using immersive audio involves the co-creation of both music and virtual space.

**Technical Research**

The physiological and psychoacoustic underpinnings of human spatial hearing are described by Wenzel et al. (2017). Lee (2014, 2020) describes the perceptual differences caused by various surround recording techniques involving height information and their implications for recording applications. Zotter and Frank (2019) offer a thorough exploration of the ambisonics format, with suggestions for recording and reproduction methods along with an
explanation of their mathematical underpinnings. Rumsey (2017) surveys the various surround sound formats that have developed over time and discusses models for evaluating the perception of surround sound quality.

Leonard (2017) describes several practical methods and considerations for mix engineers working with the creative aspects of immersive audio. Some examples include the creation of depth between sound sources occupying a similar area by routing some sounds to speakers and others to phantom locations, thereby moving the phantom-imaged sources perceptually further back in the mix. Leonard also suggests a combination of delay panning with low-pass or high shelf filtering to create time-of-arrival effects that enhance the localization of sound sources even if the listener moves their head. In discussing strategies for height panning, Leonard emphasizes the spectral quality of height perception and the anecdotal experience of mix engineers who have found that higher-frequency sounds are more effectively panned high than low-frequency sounds. Furthermore, Leonard suggests that height effects may be enhanced by boosting content above 7000 Hz.

Unpublished sources of information include the Immersive Audio Academy web seminars hosted by the Audio Engineering Society (AES). These events have included presentations by immersive audio software manufacturers who explain the background of their products and offer practical demonstrations of how to produce and deliver content. Presenters have also showcased use-cases for immersive audio technologies in various production scenarios such as live sound and broadcast. To a lesser extent, they have also featured discussions with mix engineers and creatives who have shared their experiences working with immersive audio.

While all the above-mentioned sources have been helpful for understanding and working with immersive audio on a practical level, I have still felt at a loss for understanding how to connect them to the creative aspects of music production. My experience was that the technical research answered the question of how immersive audio works, but not why I should work with it. Hence the title of my thesis, “Lost in Space”, referring to my resulting artistic confusion by being almost overwhelmed by the potential of immersive audio technologies, without understanding how it could be applied to my own artistic practice.

Thus, I became interested in how music can be conceived from the start for immersive audio playback, even to the extent that it is dependent on an immersive audio format for its full
artistic effect. To explore that question, I became interested in how composers have historically worked with acoustic spaces and exploited spatial effects in their compositions.

**Historical Approaches to Sound in Acoustic Spaces**

The history of music is interwoven with the acoustic spaces for which it was composed. Boren (2017) traces a history of the relationship between acoustics and music beginning with archaeological traces suggesting the use of reverberant caves for music performances by prehistoric humans and ending with the spatially informed music of 20th century composer Henry Brant.

Boren discusses a hypothesis that the development of polyphony resulted from the lingering of notes in the reverberant halls of proto-Romanesque churches, and states that if true, then it would “…indicate that the immersive component of sound has affected not only our spatial arrangements of music but also the very fabric of how music theory itself has developed.” (2017, p.43).

An example of music that explicitly relies on the use of space is found in the *coro spezzato* music of multiple separated choirs that emerged during the Venetian renaissance. Boren makes a connection between music production and *coro spezzato* music, when he compares the compositional practices of Adrian Willaert (the first major composer of *coro spezzatto* music) to the practices of mix engineers trying to balance a stereo recording. In making the comparison, Boren points to Willaert’s inclusion of a wide spectral range in both choirs to balance the experience of listeners situated closer to either choir as well as his use of a double bass line for cohesion.

Composers have recognized the importance of space for their music. Composers such as Bach and Wagner composed explicitly for specific venues, with Wagner going so far as to require the building of the theater at Bayreuth to realize his artistic vision (Blesser and Salter, 2006). Berlioz even saw the acoustic space itself as an instrument:

> Many fail to recognize that the very building in which music is made is itself a musical instrument, that it is to the performers what the sound board is to the strings of the violin, viola, cello, bass, harp, and piano stretched above. (Bloom, 1998, as cited by Boren, 2017)

Spatial aspects of music composition have also been used to aid storytelling, as can be seen in the staging of ensembles by Charles Ives in *The Unanswered Question* (1908), where the
separation between onstage woodwinds and trumpets and offstage strings represent a
metaphysical separation between thinkers asking existential questions and the silence of the
cosmos (Boren, 2017).

**Immersive Music Production at the Intersection of Acoustic Practice and Virtual Spaces**

The ways in which composers have historically utilized acoustic space has been contingent on
the affordances of the physical space itself. However, music producers with access to
immersive audio technologies can create three dimensional, virtual spaces for their music.
Thus, rather than tailoring music to a pre-existing physical space, music can be created
through a dialog where the virtual space informs the compositional process and vice versa.

It is in this realm, at the intersection between the virtual spaces made possible by immersive
audio technologies and the ways in which composers have worked with acoustic space in the
past, where I consider my work to be situated. This implies a kind of co-creation between
music and space as described by Blesser and Salter:

> …the introduction of virtual spaces allows music and space to be designed simultaneously in a
> truly two-way relationship. The real revolution in musical space may, in fact, be that space
> becomes a real-time artistic activity. Becoming an architect of music space thus becomes possible,
> as well as important, unlike previous periods where space was a building that lasted centuries.”
> (2006, para. 10).

It is my intention with the following three case studies to contribute to the field of immersive
audio by sharing an account from the perspective of a music producer involved with
composing for virtual spaces, and how this work has changed my outlook on the composing,
recording and mixing processes.

**Ethical Considerations**

This work has been completed in consideration of the ethics guidelines put forth by the
Swedish Research Council (VR, 2017). Therefore, I have requested permission from all study
participants to include their names and excerpts of their performances.
Three Case Studies in Music Production using Immersive Audio

In the following section I present three cases studies in music production using immersive audio and describe how they have affected my artistic practice. To the extent possible, I have highlighted aspects of the creative and production processes that I feel are common to immersive audio in general. This includes questions about compositional approaches for virtual spaces as well as practical considerations for recording and mixing.

Case Study No. 1: 360-video Concert at Saint James's Church

The conception of this project began toward the end of 2020, when virtual concerts were a necessity due to the Covid-19 pandemic. While virtual concerts provided an opportunity for audiences to experience concerts remotely, I became interested in how VR technology could make those experiences more immersive.

I was particularly interested in the idea that the listener would have the ability to interact with their environment, at least to the extent that they could shift their attention to different sound sources and have the auditory perspective change in real-time. The idea behind this kind of interactivity was to give agency to audiences, rather than force them to only listen to a predetermined mix.

To give the audience the ability to interact in real-time with the listening and viewing perspective, I decided to work on a 360-video. 360-videos allow for three degrees of movement along the x (front-back), y (left-right), and z (up-down) axes. Rotation along the x axis is referred to as roll, rotation along the y axis as pitch, rotation along the z axis as yaw, and collectively these movements are referred to as three degrees of freedom, or 3DOF (Roginska, 2017). These head movements can either be tracked by a VR headset or simulated by changing the perspective when viewing the video on a two-dimensional screen.

My intention with increasing the audience’s immersion was to explore ways to create more impactful virtual experiences, since I often felt quite disconnected from the virtual events that were commonplace during the pandemic.

Another effect of the pandemic was that public spaces were for the most part empty, and it was in this context that I was able to collaborate with Saint James’s Church in Stockholm for
the recording session. In many ways the location of Saint James’s Church, as a church located in central Stockholm, resonated with the themes of the project. Churches are, for many, a place of comfort and solace, and a space that brings people together. Since the pandemic had limited the possibilities of visiting Saint James’s church, I felt that a VR experience could at least open the doors for a virtual visit where one could experience not just the sights but also the sounds of the space. Furthermore, the interactivity afforded by VR allows listeners to go back and experience the live performance in infinite variations depending on how one chooses to engage with the performance.

To create such a concert experience and make it available on a readily available distribution platform, I collaborated with the production company IVAR Studios, to create a 360-video that was then uploaded to YouTube, since YouTube supports the playback of both 360-videos and ambisonics audio.

To connect the musical expression to the immersive audio techniques that I would be working with, I composed the music for a string quartet surrounding the listener, so that the listener would be encouraged to turn around to experience both the space and the immersive effects in the music.

Method and Results

The piece consists of three movements that were recorded in ambisonics and 360-video at St James’s Church in Stockholm. I staged the listening experience with the musicians in a square formation, with the listener situated in the middle with the musicians to the front, left, right and back (figure 1).
My goal was to develop my compositional technique by incorporating immersive effects in the conception of the piece, thereby creating a piece of music that would be dependent on immersive audio for its full effect. Knowing that the listener would be able to change the perspective in real-time, my intention was to heighten the immersive effect of techniques such as call and response and imitation, by encouraging the listener to change their perspective to follow the action of the musicians.

During the sketching phase of composing the piece, I experimented with notational methods for the x, y, and z axes and imagined how the melodic lines and phrases would interact and surround the listener from the vantage point of being surrounded by the musicians (figure 2).
Figure 2

Experiment with notational methods for X, Y and Z coordinates of sound sources over time (pg. 1).
Figure 3

Experiment with notational methods for X, Y and Z coordinates of sound sources over time (pg. 2).
In the notational method that I developed, a pan pot (O) functioned as a clef to indicate the positioning of sound sources. The range of the immersive notation was restricted to the second leger line in both directions, thus representing the extremes of positional information.

The x axis was represented by an open notehead with an x inside of it, the y axis by an open notehead, and the z axis by a breve (figure 4). Rather than specifying rhythmic values to the representation of the axes, I instead used figures to show their panning over time, with the z axis using a solid line (—), the x axis a dashed line (---), and the y axis a dotted line (•••). Thus, the positional information of the instruments only needed to be restated at the start of each measure, with the lines filling in the changes over time. If a more precise rhythmic definition were required, the notational system could have been augmented to show subdivisions.

Figure 4

Musical symbols representing x, y, and z axes.

∞ = X-axis (front-back)
○ = Y-axis (left-right)
★★ = Z-axis (up-down)

The system worked as follows: ascending note values using in the x axis notehead represented movement away to the front of the listener, and descending note values movement away to the back of the listener. Ascending note values using the y axis notehead represented movement to the left, and descending note values movement to the right. Ascending note values using the z axis notehead represented movement up, and descending note values movement down.

By incorporating the immersive element into my compositional process, I noticed that I slowly became able to imagine the placement of the sound sources in my mind’s ear. Eventually, I felt that I could imagine the immersive information sufficiently well that I did not need to utilize the immersive notation method. Also, as it became clear that I would be recording live video where the musicians would be stationary, notating panning changes was not necessary for this project. However, the exercise was helpful in raising my awareness of
how the immersive aspect can be audiated at an early stage, and how panning changes could be incorporated at the thematic level, insights which proved helpful throughout the compositional process.

For the recording session, a Røde NT-SF1 ambisonic microphone was placed in the middle of the intended listening area with a 360-camera placed above it. The NT-SF1 microphone was utilized to capture a 360-degree listening position that was as closely aligned as possible to the 360-camera. DPA 4099 clip-on microphones were used both because a clean pickup was needed for the second case study, but also because they were more discreet and allowed for an uninterrupted view of the musicians. Additional microphones were placed in the church to capture reverb and ambiance. The goal of this miking setup was to have the NT-SF1 provide the overall listening perspective and the spot mics to increase the localization of the sound sources.

The recorded audio tracks were combined in the DAW Reaper and spatialized using a series of tools made by the company Dear Reality. The 3D panning was made using a plugin called DearVR Pro. Each instance of the DearVR Pro plugin was connected via Open Sound Control (OSC) to another computer running the software tool Spatial Connect, which takes the 3D panning information in the DearVR Pro plugins and projects it onto a virtual room that then can be viewed with a VR headset. Rather than working in a generic virtual environment, I loaded the 360-video to Spatial Connect and was thus able to pan and automate the sound sources in a VR rendition of the video recording made at Saint James’s church. All the audio tracks sent the spatialized audio to a master print track which in turn sent the audio to a track containing the DearVR Ambi Micro plugin, which allows for binaural rendering of the immersive information from each individual track as well as live head tracking of the VR headset.

Besides allowing for a more inspiring working environment than the usual 2D representation of audio I am accustomed to when working in a DAW, the VR experience allowed me to follow the movements of the musicians when automating the position of the spot mics. This added a sense of dynamic movement to the spot mics that was reinforced by the visual effect of seeing the musicians moving. Importantly, it was the virtual environment itself that gave me the idea to pan in such a way, because the change in perspective made it both an easy and intuitive gesture. Achieving a similar effect by drawing automation with a pencil tool would have been extremely time-consuming and would likely not have been as accurate.
Case Study No. 1 resulted in an interactive 360-video with immersive binaural audio (Case Study No. 1). The interactive element resulted from the viewer's ability to manipulate the visual and auditory perspective in real-time. This can be done either by manipulating the perspective by hand when viewing it on a two-dimensional screen, or by rotating one’s head when viewing it with a VR headset.

Analysis

Working with Case Study No. 1 marked my first experience producing music in an immersive context and heightened my appreciation for the role of space at all stages of production.

Working with spatialization during the compositional stage has led me to audiate the location of the sounds sources when conceiving musical materials. This is an aspect of audiation that I had not worked with before this project and which came about due to my desire to create music that required immersive playback for its effect.

Previously when I would transcribe music that I would hear in my head, I would not perceive the sound as coming from a specific direction or as being situated in a specific environment. Now when I compose or sketch out ideas, I ask myself where the sounds are placed, where I am situated in relation to the sound, and what kind of room I am in. I find that this process increases my creativity, as I can employ other senses to visualize for example the size of the room, the materials of the walls, and seeing the performers in my mind’s eye. Thus, a sense of immersion is at play from the start and informs the compositional process.

By actively considering space at the compositional stage, in many ways my recording and mixing approach are clarified as I have a sense of direction for balance and placement (even if new ideas come up along the way). Therefore, I more carefully consider how the interaction between the choice of recording location, placement of the sound source in the room, and placement of the microphone(s) affect the sense of location and perspective that I am looking to achieve. This allows me to make choices upfront that get me closer to the end goal, rather than say, close miking every instrument in acoustically dry conditions using the cleanest signal path possible, only to have to labor during the mixing stage to recreate the sense of room and perspective. As a straightforward example, if I know I am going to want a timpani to sound far away, it is far easier to place the timpani in an appropriately large room and emphasize a pair of room mics over spot mics.
While this may be seen as an obvious conclusion, my point is that considering the immersive effect and placement from the outset enables me to make these kinds of decisions upfront more confidently without having to worry about whether I will change my mind later. While in the past I may have made miking decisions from a more technically minded standpoint of wanting to get the cleanest possible sound to have more options during the mixing stage, working with immersive audio has made it so my miking decisions are more driven by artistic considerations.

This has the benefit of giving me a starting point in the mixing stage that is closer to what I want to achieve sonically in the end. In my earlier approaches to mixing, I would work each track one by one and essentially try to find the goal I was working toward, rather than having a goal to begin with. By having a perspective in mind from the outset, I can make mixing decisions that are informed by an end goal in mind, with the consequence that I now complete my mixes faster and enjoy the process more.

Case Study No. 2: VR Concert in Nathan Milstein Hall

The 360-video provided an interactive experience by offering full freedom of head movement. However, I also wanted to create a more immersive experience by including freedom of movement of the body. This kind of freedom would provide a closer simulation of a real listening experience and free the listener to choose their own “sweet spot” by positioning themselves in an area of their choosing at any given point in the program.

Currently, VR technology offers a practical means for creating virtual spaces where listeners can move about freely. Therefore, I utilized the recordings from the 360-video to create a VR version of the same concert. This VR concert experience was placed in a 3D model of Nathan Milstein Hall at the Royal College of Music in Stockholm (KMH), where the listener has full freedom of movement of both their head and body.

This kind of freedom required the tracking of the pitch, yaw and roll of the listener's head, as well as the x, y, and z coordinates of the listener. Collectively, this kind of tracking is referred to as six degrees of freedom, or 6DOF (Roginska, 2017).

To not have a fixed sweet spot in the VR experience, the placement of the sound sources needed to be calculated in relation to a virtual environment, rather than in relation to a fixed speaker layout. Furthermore, the reflections and reverberation from the surfaces needed to be
calculated in real-time based on the position of the listener at any given time. Thus, it was necessary to create the project using a game engine rather than a traditional DAW.

By creating an interactive VR concert, my goal was to explore new kinds of music experiences for listeners. As with the 360-video, my motivation for the project was largely due to the lack of truly immersive virtual concerts during the pandemic.

Furthermore, I became intrigued by the potential of VR music experiences to provide unique benefits over live concert experiences. For example, in live concert experiences, fixed seating or physical obstacles often force listeners to stay in a specific area that provides a less than optimal listening position. In a VR experience, a listener may place themselves wherever they would like.

Besides offering a more enjoyable listening experience, this could also have implications in an educational context, where one could more easily study individual parts by moving physically closer to them. Or, composers could explore the ways their music will interact with a physical concert hall, by placing their MIDI mockup arrangements in a virtual copy of the same concert hall.

**Method and Results**

The project was made in collaboration with Albin Lindetorp, a 3D animator based in Stockholm, who created a 3D replica of the Nathan Milstein Concert at Hall at KMH. I also collaborated with Henrik Langemyr, one of my fellow master’s students in the music production program at KMH, to create 3D replicas of the musicians’ instruments. The 3D models and audio were all combined using the Unity game engine.

The spot mics and their automation movements were turned into audio objects with associated Unity animation files by bouncing the individual audio tracks and rendering their automation using the DearVR Spatial Connect plugin. This allowed the x, y, z coordinate information recorded in my DAW session to be translated into x, y, z coordinates for the animation files of the virtual instruments. Thus, it could be said that the movements of the musicians were transposed from the 360-video to the VR experience.

I only utilized the spot mics in the VR experience because the game engine calculates the reflections and reverberation of the listener in real-time, as there is no fixed sweet spot. I used the software DearVR Unity as the spatializer plugin in Unity and was thus able to choose a
The reverb characteristic that I felt most closely simulated the real acoustics of Nathan Milstein Hall.

The VR Experience features the same music and staging as the 360-video, but in a VR experience situated in a replica of the Nathan Milstein Hall at KMH. The VR experience expands on the interactive element in the 360-video by allowing the listener to walk and teleport around the hall. As such, the space is presented as an open-ended area for exploration and listening.

Analysis

The freedom of movement offered by VR also raises some considerations regarding compositional, recording, and mixing approaches.

In composing for a virtual space, one must consider what kind of sound stage one is envisioning, and where the listener should be granted access. Furthermore, it is entirely feasible for a listener to have special physical abilities, such as teleporting, flying or jumping high due to lowered gravity.

The questions I ask myself now are what level of freedom for the listener serves the purpose of the music, and how are sound sources placed in a way that the music can be enjoyable from as many vantage places as possible. One possible answer is that there is no need to make such decisions, as the point of the virtual experience is that the listener is free to choose where to listen. Another possibility is that the composer can have an intention for the piece, a kind of preferred outcome that guides the creative process that may take the form of visual aids or restrictions (perhaps borrowed from game design concepts), while knowing that the audience’s experience will ultimately be out of their hands. Either way, when composing for a VR space, it is important to consider that the listener’s behavior is less predictable.

In recording content for a VR experience, it makes sense to record individual sound sources, as the perspective of the listener will be calculated in real time making room mics unnecessary. Furthermore, mono sources are easier to manage in a virtual space, as they are more easily heard as point sources and do not suffer from phasing effects. Also, sound sources intended to be played back in a virtual space should be recorded as dry as possible, so that the early reflections and reverberation can be calculated by the game engine. Furthermore, audio objects benefit from a clean recording free of bleed (Tsingos, 2017).
Thus, if a production is intended to be listened to in a more traditional playback format and a VR experience, it is important to include close miking in the recording process. Depending on the type of room and position of the musicians, it might be necessary to utilize clip-on mics, or even contact mics and DI's for minimal interaction with the room.

From the standpoint of a mix engineer, one must consider that the listener will not be subject to a fixed position and may alter the balance, panorama, and spectral content simply by moving around. Mix engineers are thus forced to abandon a certain amount of control, and thus the question of how to balance can become more complicated. On the other hand, by not forcing the listener into a fixed listening position, mix engineers are also freed from having to make decisions from a limited vantage point, and arguably can evaluate the quality of their mix by exploring it in the three-dimensional space during the mixing stage. Thus, mix engineers in VR space are creating potentialities for unlimited listening experiences, and only need to ensure that the balance and panorama of the sound sources are appropriate for the space in a general sense.

**Case Study No. 3: Dolby Atmos Music Project**

The first two case studies were geared towards 360-video and VR technologies, which normally use ambisonics for immersive audio. Another form of immersive audio can be found in the multi-channel format Dolby Atmos, which has evolved from the traditional 5.1 and 7.1 systems found in cinemas and home theaters.

Dolby Atmos offers a combination of traditional multichannel sound sources called ‘beds’, as well as individual sound sources called ‘objects’, whose relative position can be scaled to any room and playback configuration using embedded metadata.

While Dolby Atmos first emerged as a surround format for film production, it has in recent years become used in music production, especially after Apple Music began supporting Dolby Atmos in the summer of 2021.

While the scalability of Atmos is one factor that makes it appealing for music production, perhaps a more important aspect is the possibility of audiences being able to enjoy a surround music mix through binaural rendering on their headphones. Thus, there is a viable market for distributing Atmos mixes to the public since consumers do not need to install expensive speaker systems.
Many producers and mix engineers are now creating both stereo and Atmos mixes of their music, and there is much interest in finding ways to work with the format. While many mixers are upmixing existing catalogues, my interest is in finding out how working with immersive audio from the start can lead to new insights in the music production process. Therefore, Case Study No. 3 consists of a work that was conceived for immersive playback, where the initial musical ideas were immediately spatialized, and the overall approach was informed by the possibilities of using Dolby Atmos.

Method and Results

For this project, I explored music production with a conception of 3D space at the outset by utilizing objects in Dolby Atmos. The inspiration for the piece came from multitracking and layering individual notes using an electric guitar, which made me want to explore the process of overdubbing in an immersive audio context.

Given the much larger soundstage in which to place the sounds, I wanted to see how far I could push the concept of overdubs while still maintaining some sense of clarity in the mix. To achieve a dense texture, I decided to record a few separate musical lines in multiple octaves, which could then be stacked vertically by increasing frequency, i.e., lower frequencies at the bottom and higher frequencies at the top.

After recording the guitar parts, I created a short loop that I played back for two hours while recording a series of improvised lines on a piano. I recorded each line in several octaves.

The piano was miked with three microphones by the open lid of the piano, one by the hammers, one in the middle, and one aimed at the bass strings from the side. These microphones were intended to accent the various octave doublings. For example, I intended the mic placed by the hammers to be used for the higher octave doublings of the lines, the middle microphone for middle register octaves, and the bass microphone for bass register octaves. Thus, I could choose from different microphones for different takes in post-production without having to stop the flow of the improvisation session. I also placed one microphone in the room about 2 meters away for a more distant sound that I could blend in.

I then edited the various lines I had come up with during the improvisation session and composed the piece by reordering the various loops and motifs. I layered the overdubs
vertically by increasing frequency and blended the piano mics according to their function as described previously.

I then arranged string and percussion parts that were primarily doublings of the piano parts that I had improvised. The string recording consisted of a Rode NT-SF1 ambisonics microphone that was decoded to an Atmos bed track using the SoundField by Rode plugin, as well as spot mics. The percussion tracks were recorded through a combination of spot mics and a pair of stereo room mics that were aimed at the back of the recording room for use as a rear surround ambience track.

While all the various parts were recorded to a click, I was careful to quantize as little as possible, and in most cases I did not quantize at all. I chose to use minimal quantization because I felt the sense of immersion was greater due to rhythmic looseness in the overdubbed parts. Thus, rhythmic looseness became a tool both to soften the sense of the downbeat and to heighten the sense of immersion.

Analysis

In Case Study No. 3, I set out to work with Dolby Atmos from the start rather than upmixing the music after producing it. There are currently several challenges to this kind of approach. First, as of this writing, there are currently only three DAWs, Nuendo, Logic and Fairlight, that feature a native Dolby Renderer, which is a requirement to be able to record audio to a Dolby Atmos session. Second, running a Dolby Atmos session generally requires a high buffer size, making it a necessity to record audio using direct monitoring. Third, as the project grows, it is likely that plugin effects will need to be either frozen or rendered, making it more difficult to make on the fly adjustments.

These challenges aside, I still found it compelling to produce music in the Atmos environment from the start, because I could evaluate the relationship between the compositional idea and its immersive effect quickly. In doing so, I found inspiration for new musical ideas that I likely would not have conceived of had I not been able to test its immersive effect. In other words, the spatial effect was intertwined with the development of the musical materials.

This has implications for my compositional language not just in terms of producing in Dolby Atmos, but also when composing for the stage. If I were to compose a piece for a large ensemble after working with Dolby Atmos, I would take far more consideration of how the
various musical ideas interact both within the ensemble and within the space. This would affect not just how the lines are distributed among the instruments, but also the seating arrangement of the musicians.

In the recording stage, working with Case Study No. 3 has highlighted an important correlation between recording for ambisonics projects and for Dolby Atmos. By utilizing ambisonics microphones as the main recording array for surround recordings, it is possible to create recordings that are compatible both in the 360, and Dolby Atmos domains. Furthermore, recordings using ambisonics microphones allow for rotating the listeners’ orientation and narrowing or widening the stereo image.

Thus, I have come to realize that a universally applicable recording format for immersive audio productions could consist of a central ambisonics array combined with mono spot mics. The ambisonics array would provide maximum compatibility with various output formats and flexibility in the post-production stage. The mono spot mics in turn, make it possible to enhance the spatial effect of 360-videos, create audio objects in Dolby Atmos, and audio sources in VR experiences.

On the mixing side, working with Dolby Atmos has highlighted the possibilities and challenges of working with the height parameter. Whenever an object is raised or lowered, its spectral information changes quite drastically owing to how we hear height, with lower frequencies being reduced the higher an object goes. Furthermore, I noticed that increasing the height of sound sources caused a perceptual lowering of its volume. Thus, moving an object in 3D space is akin to making fader and EQ adjustments at the same time.

The challenges I have faced with this process have been trying to find a way to combine the kinds of balancing and equalizing I am used to doing, with the spectral effects of panning objects. The conclusion I have reached at this point is that it is easier for me to pan the objects in 3D space first, and then apply further EQ and volume adjustments. This is because the effect of the spatial placement is so significant that I need to hear those effects before I can make informed EQ and balance decisions.

One benefit of the spectral effects of object-based mixing with height, is that many mixing problems can be solved simply by adjusting the height parameter. Associated spectral and volume change in many ways takes care of the problems of sound sources competing for the same frequency range. When doing so, it is important to check how well the mix translates to
a stereo downmix of the same material. If there is no accounting for the loss of spectral information in the stereo downmix, then masking problems may occur.

Besides offering solutions to technical mixing problems, the height parameter also allows for new kinds of artistic decisions. In Case Study No. 3, there is a string melody that enters towards the end of the piece that I wanted to sound like it was coming from a different space altogether and offer something akin to a commentary on the rest of the music. By raising the string melody high and sending more reverb from it, I was able to clear a new auditory space as the strings were separated from the ground level. Importantly, it was having access to the height parameter that influenced me to make this artistic decision.

Final Reflection

I have presented three case studies that each highlight different approaches to music production with immersive audio and how this work has affected my practice as a music producer and led to new insights in my views on composing, recording, and mixing.

In the first case study, I discussed how working with a 360-video led to an increased sensitivity to the localization of sounds when composing and audiating music, as well as helping me make more informed recording and mixing decisions due to a clearer understanding of the intended soundstage.

In the second case study, I looked at how composing for VR spaces has raised new questions for me about how to compose for virtual spaces when the listener has full freedom of movement, as well as highlighted the importance of recording clean close miked recordings to allow the game engine to render a virtual listening environment more efficiently. I also suggested that mix engineers will need to relinquish some amount of control over the listening experience and consider a shift in approach toward creating potentialities for many different listening scenarios.

In the third case study, I argued that while producing in Dolby Atmos from the start poses many challenges, it is still a favorable approach for creating works that are intended for spatial playback. I also discussed how a combination of an ambisonics microphone and mono spot mics can be seen as universal recording format that translates across the various
immersive audio technologies currently available. Finally, I discussed my observations of new technical and artistic possibilities when mixing in Dolby Atmos.

By highlighting the changes in my own artistic practice as a music producer when working with these three case studies, it is my hope to contribute to the field of immersive music production.

**Future Studies**

While working on the three case studies, many new questions, and possibilities outside the scope of the present work became apparent to me. The immersive nature of mixing in a VR environment made me wonder how an audiovisual approach to recording may be of benefit in other contexts. Furthermore, as I became more attuned to the spectral qualities of three-dimensional sound, I became interested in how those qualities could be introduced to stereo productions.

In the following section, I suggest future studies into audiovisual immersive production and stereo production through the lens of immersivity.

**Audiovisual Immersive Production**

In future studies, I would like to explore the possibilities of including the visual element in recording sessions. I am specifically referring to the use of 360-cameras to capture the performances of the musicians in the room. This would allow the possibility of mixing the audio in a VR space as in case study no.1, even if the video itself is never intended for release. It would also allow the creation of a regular 2D video with options to change the perspective in post-production. Also, including a video recording provides a valuable documentation tool and opens for other uses such as live streaming.

Furthermore, incorporating a 360-video in an immersive recording facilitates an audiovisual method that captures a full audiovisual representation of the recorded event that can be infinitely varied in the post-production stage. Such an approach could prove to be useful for both futureproofing one’s work as new immersive audio technologies continue to evolve, and to allow for new creative uses.
As an example of this approach, I recorded a student choir from KMH at Kronprinsesse Märthas Kirke in Stockholm. The ensemble was conducted by Mie Frederikke Bauer, a master’s student in the choral conducting program.

The choir was recorded using a single ambisonics microphone, a Sennheiser Ambeo VR mic, and a 360 camera, an Insta 360 One X2. By recording the audio and video in 360, I was able to render four different versions of the same recording: stereo (Future Studies_Excerpt 1_Stereo), binaural (Future Studies_Excerpt 1_Binaural), VR (Future Studies_Excerpt 1_VR) and Dolby Atmos (Future Studies_Excerpt 1_Atomos).

The included audio examples include the raw audio of the microphone to showcase its effect. In a full production scenario, this kind of approach could be supplemented with additional spaced pair microphones, outrigger microphones, and spot mics, etc.

An Immersive Approach to Stereo Mixing

Working with spatial audio has also had a profound impact on my approach to stereo mixing. As I became more accustomed to immersive audio, I wanted to experience a higher level of immersivity in my stereo mixes. Since stereo already provides a sense of directivity in the horizontal plane, I became interested in how we might recreate a sense of height and front-to-back separation. Such effects may be created by drawing on the psychoacoustic cues that are used to derive positional information.

Kim (2017) describes how the spectral quality of height perception (owing to boosts and cuts in various directional bands as well as a 1/3 octave band at 8 kHz that cause an overhead sensation), makes it possible to virtually elevate a sound even without height loudspeakers. In determining the difference between front-to-back signals, the shape of the ears aid by causing high frequencies to be attenuated when coming from the rear (Wenzel et al., 2017).

Knowing that our perception of sound depends on various psychoacoustic stimuli, I became interested in how we might use that information to create immersive effects in stereo mixes. Thus, I conducted an informal mixing experiment (Future Studies_Excerpt 02_Stereo).

The music consists of an improvised solo piano performance by Anna Wohlfarth, a master’s student in the jazz program at KMH. For the piece, Anna combined both prepared piano and separate guitar pedal effects that were recorded through a guitar amp. I used one microphone for the acoustic piano and one microphone for the guitar amp.
In planning for the mix, I wanted to try and apply the new audiovisual sensibility that had slowly emerged over the course of working with immersive audio in the case studies. Thus, before mixing I decided to draw a physical representation of the space that I imagined the music to take place in and borrowing from the visible light spectrum (lower frequencies correspond to red due to longer wavelengths and higher frequencies to blue due to shorter wavelengths), to draw how I imagined the frequencies would be dispersed (figure 5 & 6). My goal was not to make a completely accurate psychoacoustic prediction of how the auditory scene would play out in real-life, but rather a conceptual map that would inform my mixing decisions.
Figure 5
Mix visualization, top-down perspective.
Artistically, I sensed a haunting, elegiac quality in Anna’s music, that I imagined being played in a large church. I even imagined the materials of the church and candle lights. This kind of mental imaging was important and helpful for creating an emotional connection to the music and having a sense of direction, and I continued the process until I felt I could visualize a definite space and listener perspective. Thus, the idea of the music, space and immersion were correlated in my mind.

I experimented with several methods to try and recreate the acoustic space I imagined in my mind. One method was to EQ the left and right sides differently to imagine different dispersal of spectral content due to the physical characteristics of the piano. Another was to pre-delay the guitar amp to in a sense “push it back”.

I also defined four separate stereo delays and reverb, each representing front left right, surround left and right, rear surround left and right, and top surround left and right.

I conceived of the “front” delay/reverb zone as consisting of a narrower stereo width, with a boost in the high frequencies to correspond to its position as coming from in front of the listener. The surround delays/reverb zone was panned about forty-five degrees, with a longer pre-delay and reduction in high frequencies to represent reflections coming from the side walls. The rear delay/reverb zone was fully panned, with additional pre-delay and attenuation.
of high frequencies to create a sense of reflections coming from a back wall. The top left/right reverb/surround zone was panned narrow, with a reduction in low-frequencies and a boost around 8kHz to simulate a height effect.

While my experiment was informal, it nevertheless provided new inspiration for continuing to explore stereo mixing. In a sense, by working with immersive audio my thinking on what it means to mix in stereo has shifted, as I now have a goal of situating the listener in a three-dimensional space with a defined listening perspective. Furthermore, by truly imagining the audiovisual scene before mixing, I experienced a new curiosity and ease that I had not experienced with mixing before.

Thus, a future research question could involve working with immersive audio as a method for developing new approaches to stereo mixing.
References


