

Araştırma Makalesi

Etnomüzikoloji Dergisi
Ethnomusicology Journal
Yıl / Year: 6 • Sayı / Issue: 2
(2023)



Creative Panning Techniques for 3D Music Productions: PANNERBANK Project as a Case Study*

Oğuz ÖZ**
Can KARADOĞAN***

Abstract

The spatial arrangement of sounds in a three-dimensional space is a much more complex and meticulous design process than in a two-channel stereo mix. In particular, with dynamic panning methods around a fixed head, there is no certain rule for the placement of sound sources using object-type channels. The boundaries are set with the rules of psychoacoustics and hence, limited into a perspective that the listener can perceive. The software mentioned in this article is developed to create a Dolby Atmos mix designed for the audiovisual art project called MINDFLOW has taken the workflow to a further level: This work is designed by modeling a moving listener who walks through a fixed set of sound sources inside an exhibition using the original program PANNERBANK. The design and implementation of this complex dynamic panning model produced on the Max/MSP platform is discussed in detail.

Keywords: 3D Audio, Max/MSP, Dolby Atmos, Programming,

* Makale Geliş Tarihi: 10 Ekim 2023 Makale Kabul Tarihi: 24 Ekim 2023

** Produced from the doctoral thesis of the first author.

*** Research Assistant, Istanbul Technical University, Turkish Music State Conservatoire, ozo17@itu.edu.tr ORCID:0009-0003-9874-6643

*** Professor Doctor, Istanbul Technical University, Turkish Music State Conservatoire, karadoganc@itu.edu.tr ORCID:0000-0003-3611-6980

Creative Production Techniques

Üç Boyutlu Müzik Prodüksiyonları İçin Yaratıcı Panlama Teknikleri: Örnek Uygulama Çalışması Olarak PANNERBANK Projesi

Özet

Müzik teknolojisi tarihinde tasarım araçlarının geliştirilmesi, medya formatlarının iyileştirilmesi ve de sunum şekillerinin yenilenmesi sıkça karşılaşılan bir olgudur. Bazı yeniliklere sanatçılar ve mühendisler önce büyük direnç gösterse de bu yenilikler müzik prodüksiyonu üretim süreçlerini kalıcı bir şekilde değiştirmeye adaydılar. 1960'larda mono mikslardan stereo mikslere geçildiği gibi 2021 yılında dijital müzik platformlarında üç boyutlu ses formatlarından Dolby Atmos'un sinemadan gelen altyapısıyla büyük bir üretim değişimine aday olduğu görülmektedir. Bu yeni formatların sunulmasındaki ana amaç çoğunlukla yaratıcı endüstrilere yeni ifade imkânları kazandırabilmektir. Üç boyutlu bir mekânda seslerin uzaydaki yerleşimi iki kanallı bir stereo miks çalışmasındaki göre çok daha karmaşık ve titiz tasarım gerektiren bir işlemdir. Özellikle sabit bir kafanın etrafında dinamik panlama yöntemleriyle ses kaynaklarının nesne tipi kanallar ile yerleşiminin belli bir kuralı olmadığı gibi dinleyicinin algılayabileceği bir perspektifte sunulmasına gayret edilmektedir. Yazıda bahsedilen MINDFLOW projesi için tasarlanan Dolby Atmos miks çalışması için geliştirilen yazılım bu bakış açısını bir adım öteye taşımıştır. Bu çalışmada PANNERBANK yazılımı üretilmiş, bir serginin içindeki sabit yerleşimli ses kaynaklarının arasından yapılan bir yürüyüşün hareketli dinleyicinin modellenmesi ile tasarlanmıştır. Max/MSP platformunda üretilen bu karmaşık dinamik panlama modellenmesinin tasarımı ve uygulaması ayrıntılı bir şekilde ele alınmıştır.

Anahtar Kelimeler: 3B Ses, Max/MSP, Dolby Atmos, Programlama, Yaratıcı Yapım Teknikleri

INTRODUCTION

3D audio has unique features for music producing. In this paper, the application which was designed for the extended usage of these features will be explained. The work will start with a brief history of 3D audio to understand the culture and usage of 3-dimensional audio in acoustic music and productions with examples. Afterwards, the aesthetic approaches in the music projects which were mixed with the application will be explained. The application and the technical details will be explained in detail in Chapter 3. Deficiencies and the further development will be discussed in conclusion.

Brief History of Spatial Audio Culture

Even though 3-dimensional audio production technology appeared in the late 20th century, as a concept it dates to earlier ages of music history. Before music recording and production technologies were invented, the 3D sound placement concept was used acoustically by composers. Separated choir and orchestra were used in “Coro spezzato” style in the Renaissance period in Italy. It was not only separating choirs but also composing the music in a complex call-and-response style. 2 or 3 choirs, each had 4-part harmony, were placed far apart, and sung one after another or together (depending on the music). In England, Thomas Tallis’s “Spem in Alium” was composed for 8 separate choirs of five voices each. Although the details about the original premier are yet unknown, the piece was performed with 8 choirs arranged in a circle and the audiences placed in center in 1958. British composer Vaughan-Williams wrote “Fantasia on a theme by Tallis” for three different sized orchestras. Even in the classical era the awareness of spatial sound was on the minds of the well-known two: Mozart composed “Don Giovanni” for 3 orchestras, one in the pit, one onstage and one backstage. Haydn composed “Das Echo” for string sextet for 3 trios spatially separated. In the romantic period, Berlioz used separated orchestras or instruments in his pieces such as “Symphonie Fantastique”. In the 20th century, acoustically separated orchestras or sound systems used to create spatialization by composers’ such as Charles Ives, Henry Brant, Pierre Schaffer, Pierre Henry etc. Pierre Schaffer and Pierre Henry created a tetrahedral four-loudspeaker system for their musique concrete. In 1958 World’s Fair, Edgar Varese’s piece “Poeme Electronique” was presented over 425 loudspeakers in Philips Pavilion (Boren, 2018).

By the 1960’s, multichannel audio systems for the commercial market had started to be developed. In 1969, the first quadraphonic broadcast took place. Two different FM stations broadcasted the Boston Symphony Orchestra simultaneously front and back channels in stereo format. Two stereo systems were required to experience that quad broadcast. Also in 1969, 4 channel open-reel types were put on the market by “The Vanguard Recording Society”. However, the high selling prices and the limited number of machines to play them back pushed the industry to develop the new methods. Peter Scheiber is one of the first people who works on the quadraphonic matrix

systems which compresses four channels of audio data to a single stereo file and reproduces it to a quadraphonic speaker array by employing channel separation and phase artifacts (Barry, 1975). Many others worked on surround formats like ambisonics and wavefield synthesis for not only the market but also within academia.

Binaural formats were also considered for the market since they are the simplest 3D audio technology and easy to distribute. There are productions in which transaural binaural methods were used to broadcast like CBS's *The Twilight Zone* (1985). Even though binaural was used in several albums, that could not be an industrial standard.

Popular music is only marginally touched by binaural technologies. Examples can be found in Tchad Blake producer discography (Pearl Jam, Latin playboys, Tom Waits) or in albums like Pink Floyd "The Final Cut", Michael Jackson "History", or in some Stevie Wonder and Lionel Ritchie albums, that used the holophonic technique (by Hugo Zuccarelli). Nevertheless, the application of binaural techniques in popular music stays anecdotic. This lack of success seems to be due both to technical reasons and more general considerations (Fontana, 2007).

Even though there were many engineers to develop the technology and many others wanted to use and experience 3D audio, the lack of a compact format for multichannel audio made the technology fail commercially. In the late 1990's, SACD by Sony/Philips and DVD Audio from JVC was developed. Even though these formats were used in several albums they could not be a standard production approach (Boren, 2018).

The experience and use of 3D sound in human culture has always been tied to the technological capabilities of the generation hearing it: for most of human history, this technology was limited to architectural spaces and music composition. Since the 19th century, more advanced technologies have allowed both the more accurate representation of real-world soundscapes as well as sonic spaces that have no correlation in physical reality. The former trend is now being used in the rapid development of audio for virtual reality applications, which use head-mounted displays and headphones and are thus reliant on binaural reproduction methods (Begault, 2000; Xie, 2013). The latter trend of exploring new non-physical spatial audio has been manifested in augmented reality systems, which use spatial auditory content to represent information beyond that encountered in the real world. Though these systems may use multichannel or wave field loudspeaker methods within certain controlled environments (Boren et al., 2014), they also require binaural reproduction for deployment in day-to-day life, especially given the ubiquity of in-ear headphones or 'earbuds'

in modern society (Sundareswaran et al., 2003). Multichannel methods persist for film audio and home theater systems, but in the future these markets may also face increased competition from binaural or transaural content as HRTF individualization methods improve (Boren, 2018).

Technical developments like internet streaming and the streaming speed in recently, there are new approaches for mixing and delivering the 3D audio. With the transition to the digital cinema, engineers proposed the idea of transferring the original objects in monophonic format to be rendered in playback devices. According to this Tsingos, this approach has several advantages in production and delivery, such as;

Enhanced immersion, adding height and flexible rendering across speaker layouts and environments;

Enhanced personalization, allowing consumers to tailor the content to their preferences;

Enhanced adaptability, ensuring that content is optimized across a wider range of playback devices;

Enhanced accessibility, with improved multiple language support, improved video description and dialogue enhancement;

Efficient production workflows and future-proofing of content, by deriving current or future deliverables from a single object-based master mix (Tsingos, 2018).

Dolby Atmos Format

Even though other companies used object-based audio format in audio production such as DTS:X and Auro3D Max, recently Dolby Atmos is used in most music production. The first Dolby Atmos content, the movie “Brave” by Pixar was released in 2012. Even though the format was focused on movies, it can be used in other fields such as music, gaming, live broadcasting, etc. Production companies started to use the format for music production since streaming platforms such as Amazon Music, Tidal and Apple Music started to support it in late 2010’s. With the help of the format’s flexibility in playback devices, technology companies are working on to use the formats for different experiences with the combination of other technologies such as head-tracking, virtual-reality, augmented-reality etc.

In any field except interactive products like gaming and VR, 3D audio is used to create an immersive space by panning the sound sources instead of the listener. However, technology allows the content creators to work on more interactive works by locating the moving listener into the center of the steady sources. This was the main purpose in the project which this paper focused on. New devices were designed to create this new aesthetic approach.

PROJECT OVERVIEW

Mindflow is a contemporary artwork designed and implemented by HA:AR, an art duo with Arda Yarkın and Hande Şekerciler. HA:AR combines diverse and contemporary tools of art and technology in its works. Similarly, in their project MINDFLOW the focal point of the work lies in dance choreography (Şekerciler & Yalkın, 2023). The dancer's movements were recorded by motion capture devices. An AI tool translated this information into artistic videographic patterns. The first step of the interactive part had classically trained musicians improvise while watching this video solo. After these takes are done, all individual recordings are placed in an exhibition hall using individual speakers. Audiences experience this interdisciplinary work by wandering between speakers playing improvised recordings and screens showing videographic images (Figure 1).

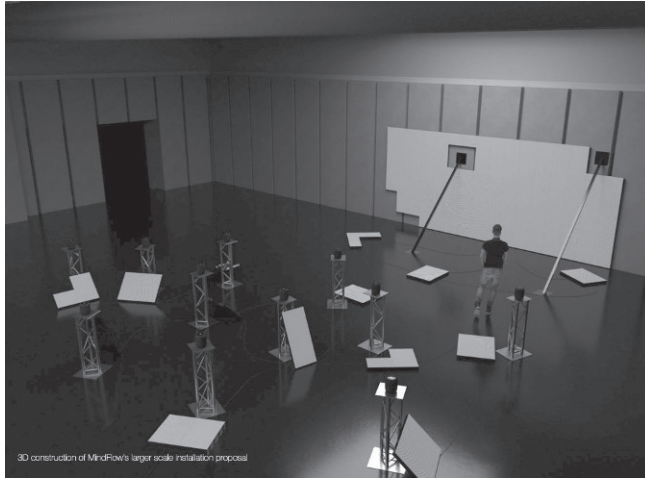


Figure 1. 3D render of large scale installation proposal from Mindflow project handbook (Şekerciler & Yalkın, 2023)

Mindflow is an immersive installation that consists of 9 videos and 9 sound sources which unifies architecture, the art of dancing, artificial intelligence and computer-based visual technology. Mindflow is aiming to break the traditionally-constructed, strict boundaries of classical music and achieve a new composition by transforming music with randomness, intuition and improvisation. While doing so, Mindflow is aiming to visualize the creation process by utilizing generative constructions (Şekerciler & Yalkın, 2023)

As it is shown in Figure 1, each instrument has its own 360° speaker and audiences can listen separately to the musician's reaction to the images. It is also possible to hear the entire orchestra with the stereo system which is located at the entrance of the venue.

The audio tracks used for the project were recorded at Babajim Studios in Istanbul. 8 different instruments were recorded for the project: harp, clarinet, cor anglais, cello, violin, double bass, piano and 2 percussions. Both close and room microphones were used to record the performances. This decision was important to create a natural sense of acoustic space in the mixes. The musical performances were not edited correctively to preserve the musicians' feelings about the images shown to them (Figure 2).



Figure 2. Footages from the recording sessions at Babajim Studios, Istanbul (Şekerçiler & Yalkın, 2023)

As mentioned in the introduction, Dolby Atmos is a flexible format to experience production in different environments. Firstly, it is important for the artists to have this option to give access to more listeners who have different listening devices. Secondly, the format is supported by many streaming services. It makes it easier to distribute the work than using other formats. It also creates other opportunities to expand the projects.

The initial plan was to have only a 3D mix. The main idea was to create a new kind of experience and a different artistic outcome from the project. However, it was necessary to create a stereo version of the mix to release it on streaming platforms. The stereo mix should contain the same musical ideas as the 3D version as possible, even if the 3D version would keep the attention.

During the production meeting, 2 different ideas for the 3D mix were considered. One was to think about musically creative ideas, as the musicians had done during

improvisation. Some channels were personally chosen by the sound engineers to be more in the foreground for various parts, as a common approach in music production. The video shown during the sessions was watched several times with the channels, and musical mixing decisions were made. Another idea for the 3D mixing was to simulate the original installation environment. The goal was to create a virtual sound space with all the automations working simultaneously to move the listener around in it.

Mixing Strategies

As mentioned above, the release of the piece was needed to fulfill all the required formats of the digital service providers upload expectation. The artists desired to create a listener's experience close to one of the exhibition spaces that would be reflected in an immersive mix that is to be consumed on a pair of headphones. They also insisted on the head-tracking feature that Apple Music platform offers. The head-tracking feature creates the illusion of "wandering" around in the musical sound space: When it is on, you can turn your head to the source that is normally panned to the sides or to the back of your head and focus to the sound as if it is in the front.

The piece is based on isolated improvisations done by eight instruments as they were watching a dance configuration. Turning on all instrument channels did not provide gestural clarity and cross-modal metaphorical moments when combined with the visuals. So strategically, a fixed time-based plan of the walk through the sources needed to be planned and this needed to be used for both musical 3D and stereo mixes. According to the plan seen in Figure 4, the listener walks through the sources and stops at a point where 3 instruments sound closer and the rest in the background every 60 seconds. Like the acoustic exhibition experience, all instruments can be heard throughout the piece but their volumes in the mix varies gradually depending on the position of the listener.

Stereo mix in Pro Tools

The stereo mix is completed in Pro Tools. Although the improvised tracks demanded a static classical music mix approach, the idea of a moving listener through fixed sound sources inspired another approach. As used in Walt Disney's 1940 animation *Fantasia*, the stereo mix made use of unnatural sounding dynamic panning techniques and volume automation moves. In today's classical music mix standards, instruments tend to be panned to a fixed position in the sound stage and kept there without huge volume changes in the automation. As a stereo mix to be compared to the dynamically varying 3D mix, it needed to have a similar extraordinary approach. Figure 3a below shows the volume automation lines of each channel, and every sub channel group of the individual instruments using the Pro Tools Studio DAW.

As mentioned above, depending on the walkthrough plan, the pan position of the listener to a given instrument would change, depending on the pre-decided stop

point. For example, the harp would begin from a hard pan left position, but then in the second half of the walk it would be panned to the right.

Musical 3D mix in Logic Pro X

The musical 3D mix used the same walkthrough plan as well. In both the stereo and the musical 3D mixes, the sources were moved around the head of the listener, attempting to create the illusion of the listener walking through these sources. In addition to the stereo mix, a little extra movement was added to the sources to create a “seeing by listening” feeling. The automation lines were created by actively performing the write function as the mix flows, which added a certain vibe of liveliness to every source providing the listener an element of fascination in an exhibition feeling. Figure 3.b shows the automation lines for 3D dynamic panning movements using the Logic Pro X DAW.

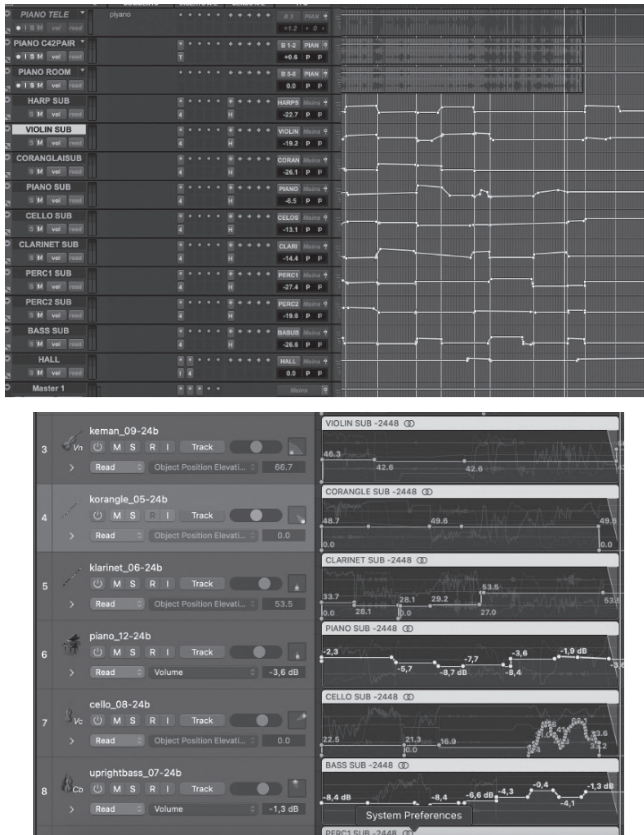


Figure 3. a) Automation lines used in the Musical Stereo mix session created using Pro Tools Studio. b) Automation lines used in the Musical 3D mix session created using Logic Pro X

Physical space simulation 3D mix

For a 3D mix that takes the physically realistic approach, we needed to use a DAW that has 2 important technical capabilities. Native Open Sound Control (OSC) support is one of the factors for communication with the Max/MSP. It should also have a customizable and assignable MIDI connection to control various parameters such as faders and pan pots. Since the main streaming platform to release is planned as Apple Music, it is important to have binaural playback of Apple Music during the mixing process. We can use almost any DAW to create the metadata Dolby Atmos needs for playback, but Logic Pro is the only DAW that allows audio engineers to hear Apple Music's binaural playback during mixing. Logic Pro was used for the realistic 3D mix because of its 2 technical capabilities.

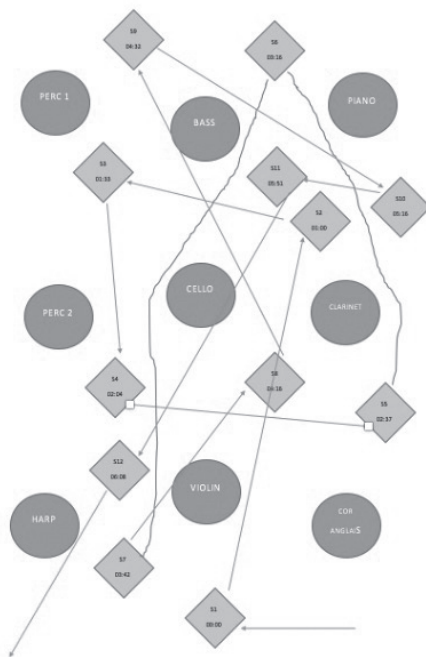


Figure 4. Placement of sources and time –based walk laps scenario that is used as a basis for all mixes

Both close and room channels were used for the entire mix. The volume controls should be set according to the intended position of the listener. When the position of the virtual listener changes in the virtual room with fixed positioned speakers, the faders of the near and room microphone channels should move accordingly. Also, the sum of the channels should be increased or decreased depending on the position of the listener. To achieve this approach, we should turn our thoughts in the opposite

direction, because technically we are panning the sources instead of the listener's position in the DAW. If we plan for the listener to move forward, we should pan the source backward. If we plan for the listener to move to the left, we should pan the source to the right. Moreover, the positions of these virtual sources should be the same throughout the piece. We planned a fixed position for all instruments as they are in the real installation. In our musical strategy, the positions looked like the picture below (Figure 4).

Headphones were used in this version of the mix, as it was mixed primarily for binaural with head tracking. The goal is to make this sense of place and movement even more realistic with the head-tracking capability offered by the Apple environment. Listeners should turn their heads according to the time code to better understand their position and interact with the mix. Listeners can focus on different music sources with the head-tracking capability.

The sound should be as natural as it occurs in the scene. We used EQ in the mix only to prevent one instrument from overpowering the others in the frequency spectrum. When the virtual listener is closer to a source, the sound becomes drier and louder than it would be in the real world. When using an EQ more drastically, the closest source will sound unnatural. For this reason, the equalizer was used in a gentler way. Compressors were also used to prevent one instrument from masking the others and clipping itself, especially if all instruments are heard equally when the virtual listener remains in the center of the room. Also, the parameters of the compressors are set to work smoothly because changing the distance makes the compressor's behavior more audible, which can be unnatural. Although we had room channels, additional reverb was used to create wholeness. Room microphones could not capture the same quality of an instrument and the room because the distance between microphones and instruments changes in the recording room. There was no automation to control musical parameters. All the artifacts and gestures created by improvisations were left as they were in the recording.

LISTENER-BASED PANNING APPLICATION FOR REALISTIC MIX

To change the paradigm and add another creative touch to this artwork, the idea of designing this new tool for listener position-based panning was accepted by HA:AR and the decision of creating separate immersive mixes was made. This article will focus on the technical implementation and artistic decision-making process of the listener-based panning approach.

Explanation of Setup and Routing in the Digital Audio Workstation

To create a more realistic approach, we used close, and room microphones together. They were both routed to an auxiliary channel to control the main volume and the limiter. Both close and room microphones were sent to an auxiliary reverb channel to automate and make the reverb sound more natural.

The sound should be as natural as it occurs in the real world. This is the reason that equalizers were used in the mix only to prevent one instrument from overpowering the others on a frequency basis. When the virtual listener is closer to a source, the sound becomes drier and louder than it would be in the real world. When using an EQ more drastically, the closest source will sound unnatural. For this reason, the equalizer was used in a gentler way. An additional EQ was also used for a reverb channel to prevent the mix sounds muddier. Compressors were also used to prevent one instrument from masking the others and clipping itself, especially if all instruments are heard equally when the virtual listener remains in the center of the room. Also, the parameters of the compressors are set to work smoothly because changing the distance makes the compressor's behavior more audible, which can be unnatural. Although we had room channels, additional reverb was used to create wholeness. Room microphones could not capture the same quality of an instrument and the room because the distance between microphones and instruments changes in the recording room. There was no automation to control musical parameters. All the artifacts and gestures created by improvisations were left as they were in the recording.

As mentioned before, the mix has an approach to create an illusion of the real world. To achieve this, the mixing engineers should think in a counterintuitive way when panning the sources, because technically the sources should be panned instead of the listeners in DAW. If it is planned to move the listener forward, the source must be panned towards backward. If it is planned to move the listener to the left, the source must be planned towards the right. Moreover, the positions of these sources should stay the same throughout the piece.

To have a convincing sonic image, all volume and pan parameters must move simultaneously. The easiest and most accurate way to have simultaneous movements was to create an application where you can control all parameters together. Creating a standalone application based on Max/MSP connected to the DAW provided the best option to achieve the intended sonic image. There are some other programming languages / programs to create an application which can be capable of doing what is aimed. Researchers of the music technology area are more experienced in Max/MSP programming, which is the reason that it was chosen.

Explanation About the Connection Between Max/MSP and DAW

MIDI (Musical Instrument Digital Interface) and OSC (Open Sound Control) are 2 options to connect between Logic Pro and Max/MSP. The plan has been changed several times due to their capabilities. The control parameters can be accessed and modified in the "Controller Assignment" window (Figure 5). The "Learn Assignment for ..." function can be used for easier access to create and modify the assignment MIDI or OSC. More details about the patches and the connection are explained further on in this paper.

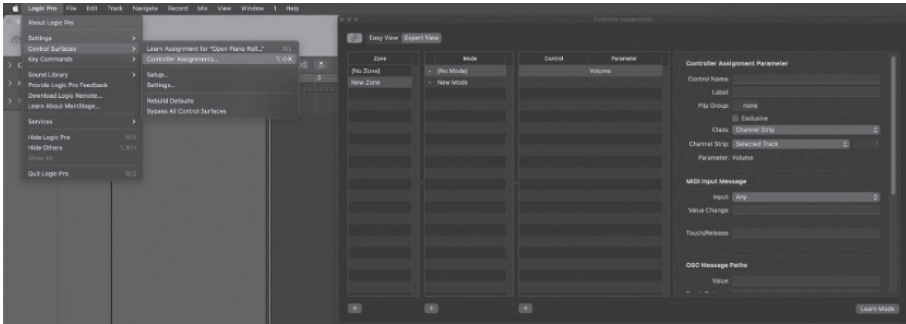


Figure 5. MIDI mapping window in Logic Pro

During patching, some problems occurred due to data flow in communication. The original plan was to send data only through the MIDI protocol. When the app was patched to use only MIDI, some parameters sometimes got stuck or did not work at all. There were no problems with individual control of volume or pan. However, data could not be properly transmitted to Logic Pro X when all parameters were working together. Also, Logic Pro X does not allow feedback via the Inter-Application Connection (IAC). MIDI Continuous Controllers (CC) signals cannot be sent via IAC, as noted on Apple’s support website (Apple Inc., 2023). Because of these problems, the OSC protocol was chosen for the connection.

At the beginning of creating the OSC setup, the main problem was that the default OSC command list (aka. layout) is only supported by controllers which are approved by Apple. One of these layouts is TouchOSC’s LogicPad. There are many parameters that can be controlled over or displayed on an iPad. However, using this layout is only possible with external iPads. Creating OSC communication between Logic Pro and Max/MSP or other applications in the same computer is not possible by default. To make this possible, Logic Pro should be tricked to see the Max patch as an iPad by setting a custom zero-configuration networking. First, the plug-in called “Zero” should be downloaded and installed to Max/MSP. After the installation, “zero.announce” object can be used to define the port, the device name as a tablet and has UDP OSC connection. Logic Pro automatically sees the max patch as a TouchOSC application (Figure 6). There were no problems when communicating via OSC. However, the researchers could not find a way to implement a new OSC parameter list. It was possible to use some standard OSC command lists, but at the time of the project there were no commands on it to control 3D panning data. The researchers tried to split parameters among different protocols with which they would work better. The MIDI protocol would be used for the parameters which cannot be controlled over OSC (e.g., 3D panning parameters) that do not need to be feedback in Max/MSP, while OSC protocol would be used for others.

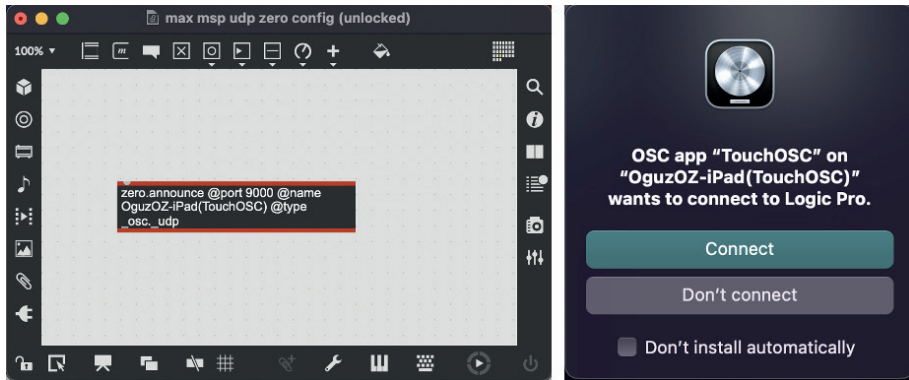


Figure 6. OSC connection between MaxMSP and Logic Pro

The performance of OSC with MIDI protocol was not proper enough to complete the mix. There were still drops in the connection while the virtual listener was moving. At that point the final decision was made to use only the MIDI protocol but with more than one virtual device. Controlled parameters were divided into 3 groups; main volume, room/close microphone mix and panning parameters. 3 different virtual MIDI ports were created on IAC Driver on Audio/MIDI Setup

(Figure 7). Each of these 3 ports was used individually to connect the groups of parameters to Logic Pro.

The final problem with the connection is moving custom MIDI assignments to the other computers. Researchers could not find an uncomplicated way to create a file which was able to be copied and pasted to other computers easily.

The only solution was copying Logic Pro's MIDI assignments file. The MIDI parameters which are assigned in Logic are stored in "~/Library/Preferences/com.apple.logic.pro.cs". The new parameters and assignments are saved in this path after each change. Copying and pasting this file without changing the name is the easiest way to move the parameters to another computer.

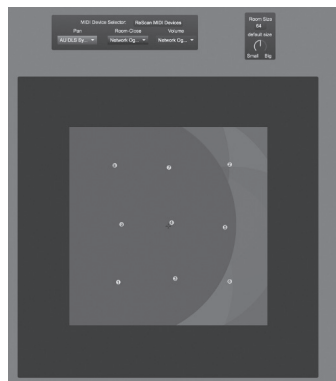


Figure 7. User interface of the Max MSP patch

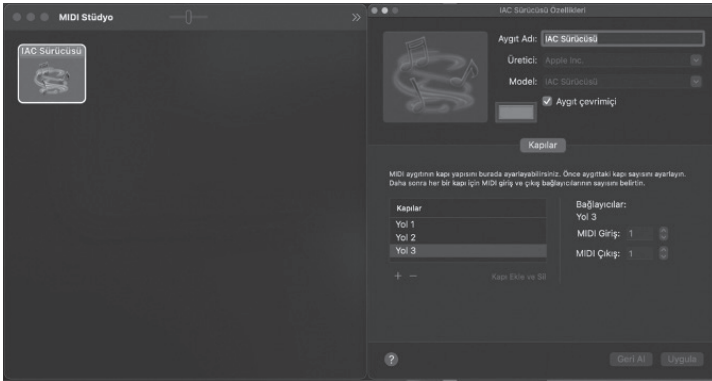


Figure 8. MIDI IAC settings

Explanation of the Max/MSP Patch

Figure 8 shows the basic user interface of the Max/MSP patch. There are 2 different main sections in the patch. The settings section is located on the upper side of the patch. In this section, MIDI connections and the virtual “room size” can be set. Users can select the different virtual MIDI devices with the pop-up menus. It is possible to scan the other possible MIDI devices by clicking the “ReScan MIDI Devices” button. Figure 9 shows the patch of the MIDI control section.

The panning section is located on the lower side of the patch. The “nodes” object is used to control the position of the listener and the sound sources. Nodes object size is controlled by room size knob.

“midiinfo” objects outputs a list of midi devices to a pop-up menu when it receives bang. A “message” object named “ReScan MIDI Devices” is connected to bang the “midiinfo” object and create the MIDI device list. “umenu” objects are connected to show and select the MIDI device. It is time consuming to select each MIDI device separately. To select devices sequentially at once, each “umenu” object is linked to each other to select the next device.

All “umenu” objects are connected to different “send” objects to send the name of the selected devices to each parameter’s dedicated “midiout” objects in abstractions. The “Room size” section is also connected to other “send” objects to send the size data to abstractions. Room size is controlling the virtual room size by changing distances between objects. The range lies within 0-127 while 64 is the default value.

The “Nodes” object is used to set the main parameters which are 3 volume faders (room, close microphones, and auxiliaries) and 2 locations information (left/right, front/back) for each instrument in the mix. The “node” represents the sound source and “slider” represents the listener. In “nodes” object it is possible to create up to 64 different nodes, however it is limited to 9 nodes because there are only 9 instruments in this project. Each “nodes” is assigned to control the locations and volume faders of

one instrument. The input of the object is used to set the visual representation of the room, to change the number of objects and to “bang” the object to get x-y axis data (Figure 10).

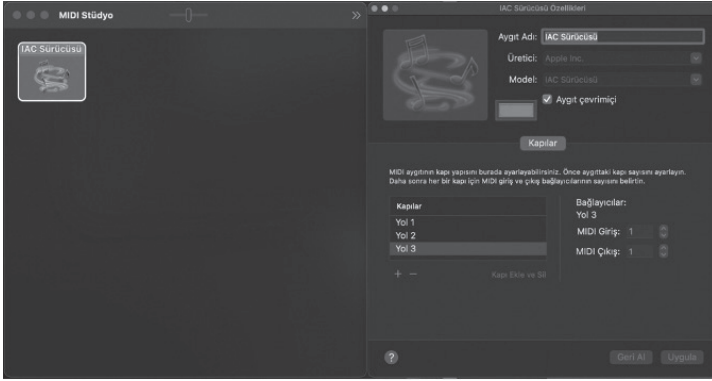


Figure 9. Patching view of the setting section

The “presentationRoomDisplay” patcher is created to control the size of the “nodes” object in the presentation view. The data obtained from the room size setting is scaled between certain numbers to make the view bigger or smaller and to make it seem to stay in the same center by changing the object’s left corner position, which is the reference point to position any object in Max/MSP. The “getxy” message box is banged continuously while positions are changing and sends the “getxy” message to send the position data of nodes and the slider from the middle outlet of the nodes object.

Nodes object gives several outputs at the same time which are optimized and assigned to different parameters.

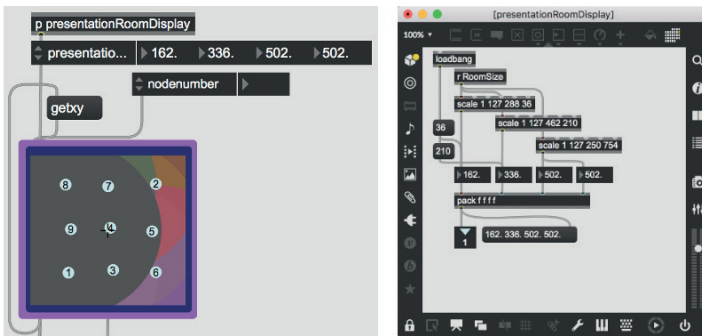


Figure 10. Patcher view of room size controller

Following outputs are used in the patch:

- Nodes position
- Slider position
- Distances between nodes and slider

Creating abstractions is the easiest and proper way to control and edit the patch in the Max/MSP environment. Changing only one abstracted patch is enough to make the same changes for all channels. Also, abstractions make the patch run like the logic of fader bank, which can give further opportunities to expand the patch. The abstraction named “ChCtrl” was created to set essential parameters to make the patch work (Figure 11). Four inputs of the “ChCtrl” object are connected to different parameters coming from the nodes object and message boxes. Three different data coming from nodes object are used for the positioning and volume faders, while message boxes set the fader bank numbers.

The message boxes named 1-9 determine the fader bank number for each ChCtrl object. The “loadbang” object bangs the message boxes to send the fader bank numbers to each ChCtrl object when the patch is opened.

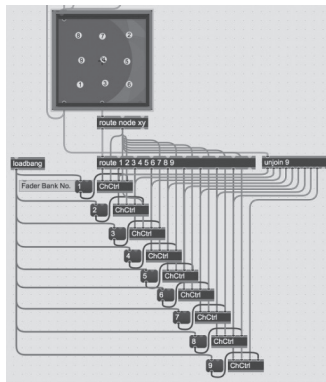


Figure 11. Connection to “ChCtrl” patch

The left outlet of the nodes object sends proximity information which shows distances between the nodes and the slider. Data is listed. Listed data should be separated, then carried to individual ChCtrl objects, which have the fader bank number the same as the node number, to control different channels. The middle outlet of the node objects is dump-out which sends several types of data. This outlet also sends the listed information which should be separated. The first route object which has “node” and “xy” arguments routes the positions of each node from the left outlet as a list and the slider position from the middle outlet. The second route object distributes the nodes position data to their individual “ChCtrl” objects.

Explanation of “ChCtrl” Object

ChCtrl object is the main abstracted patch which collects and processes data and sends them via MIDI out (Figure 12). Location information is processed on the upper side of the patch while the auxiliaries, room and close microphones volumes processing is being done in the lower side of the patch. The 8 main parts are color coded to make the explanation easier. Each input and output is named and has a green colored background. There are also 7 different receive objects to get data which is the same for each ChCtrl object (Table 1).

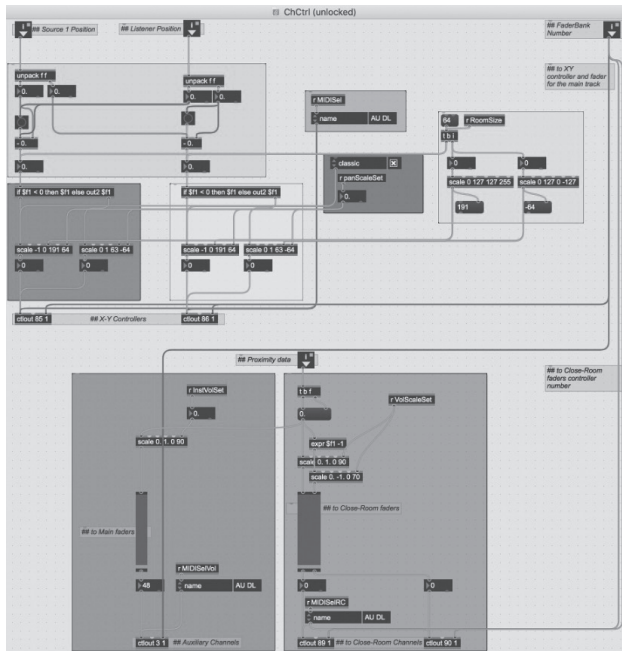


Figure 12. Inside “ChCtrl” patch

Table 1. Names of send objects and information they carry

Object Name	Area	Data Information
MIDISel	Orange	Rescan and select MIDI device for pan
RoomSize	Turquoise	Room size
PanScaleSet	Purple	Pan Law
InstVolSet	Blue	Logarithmic scaling for auxiliary
MIDISelVol	Blue	Rescan and select MIDI device for auxiliary
VolScaleSet	Pink	Logarithmic scaling for room and close microphones
MIDISelRC	Pink	Rescan and select MIDI device for room and close microphones

Green Area: The location data comes from the first and the second inlets of the ChCTRL object. They are unpacked and sent to float number objects to see the data. The subtraction object is used to calculate the location of the channels according to the virtual listener.

Red and Yellow Area: The linear data given by the nodes object are not sufficient to create a realistic soundscape while the virtual listener is walking between the sources. The movement of the sources should be changed according to the distance between them and the listener, especially when the room gets bigger. The location data, which is calculated by subtraction, is divided into 2 parts, and processed differently. If it is below zero, which means the source is positioned to the left of the listener, data is sent to the left scale object and scaled between 191 and 64. If it is above zero, data is sent to the right scale object and scaled between 63 and -64. This is the widest scaling range. It can be changed by the room size data, which is obtained at the 3rd inlets of the scale objects. Even though the panning in the DAW has 128 steps, data which is below zero or above 127 are calculated to minimize the patch work. The far-left inlets of scale objects are connected to data calculated in the purple area of the patch. The final calculated panning data is sent to “ctlout” objects, which have 2 arguments; controller number (can be changed from middle inlet) and MIDI channel (can be changed from left inlet). Controller number 85 is for the X axis and 86 for the Y axis for each channel (Table 2). MIDI channels are different for each “ChCtrl” objects. The left inlet of all “ctlout” objects are connected to their individual fader bank number which is coming from the 4th inlet of the “ChCtrl” objects.

Orange Area: The receive object carries the selected MIDI device for the “ctlout” objects which send the MIDI data assigned to control pan knobs.

Purple Area: The receive object carries the selected float number to change the exponent in the scale objects to change the scaling curve.

Turquoise Area: The receive object carries the integer room size number coming from the room size setting. The data is scaled twice. Left scale object is to control the far left and far front value. A right scale object is to control the far right and far back value. They are connected to the 4th inlet of the scale objects, which calculates the MIDI data for pan knobs and changes the scaling range.

Blue Area: The proximity data that shows the distance between nodes and slider in nodes object comes from the 3 inlets of ChCtrl object. The data is in float number format and has a range between 0. and 1. It should be scaled to the range between 0-127 for MIDI. However, in the scale object, it is scaled to 0-90 because in the DAW, the MIDI value 90 is equal to 0 dBFS on the volume fader. The “r InstVolSet” object is the receive object which carries the float value which is used to change the scaling curve to make the fader movement coherent to the pan law. The “r MIDISelVol” receive object gets the name of the selected MIDI device name which is used to transfer the MIDI signal to the DAW to control the auxiliary fader movement. The “ctlout 3 1” objects are used to send MIDI data over the controller number 3, and the same MIDI channel with other ctlout objects in the ChCtrl object.

Pink Area: The proximity data coming from the third inlet is used to control room and close microphone signals. Two separate scale objects are used to scale the proximity data differently. The one on the left is used to control the close microphone while the right one is for the room microphone. The “r VolScaleSet” object carries the float value which is used to change the scaling curve to make the fader movements coherent to the pan law. The range of the scales is different. The closed signal is scaled between 0 and 90 which is infinity to 0 dBFS in the DAW. However, 0 dBFS for room microphone sounds much more than it is needed in the mix. Scaling between 0-70 is set by researchers’ aesthetic choice during the mix. MIDI value 70 is equal to -10 dB in the DAW. The “r MIDISelRC” object receives the MIDI device name which is used to transfer the data from patch to DAW. Two separate ctlout objects are used to send scaled signals via 2 individual CC numbers and the same MIDI channel with other ctlout objects in the ChCTRL objects.

Table 2. MIDI CC numbers and data they carry

MIDI Controller Number (CC Number)	Data
85	Pan (X axis)
86	Pan (Y Axis)
3	Auxiliary channel's volume fader
89	Close microphone's volume fader
90	Room microphone's volume fader

DAW-MIDI assignments

Figure 13a shows the channel list in the Logic Pro X session. The first 9 channels are the auxiliary faders of each instrument. These channels control the main volume fader coming from each instrument's room and close microphones. Also, these auxiliary channels are responsible for 3D panning for each instrument. The stereo signal which comes from the close and room microphones does not have any 3D panning information.

They panned in stereo domain, then that stereo signal is panned on the 3D domain in the auxiliary channel. Close and room microphone channels which are used for the volume changes depending on the movement in Max/MSP are located after the 10th channel. EQ processing is also applied on these channels. To create a more realistic approach, a reverb is used in the mix. The Reverb channel is fed by the microphones' channels in post-fader routing settings. The reverb amount is not changed during the mix, but the fader movement affects the amount of dry signal which is feed the reverb. Close microphones' reverb send levels are lower than the room microphones. Even with the loudest volume level of the close microphone the signal should not have too much reverb because the louder close microphone signal would give the impression that the listener is closer to the source. For this feature, Logic Pro's default reverb plugin is used.

All the MIDI CC signals should be assigned to correct parameters in Logic Pro. The fader bank-wise design of the MaxMSP patch allows the same MIDI CC information to be used for each channel with just a different fader bank number. Figure 13b shows the MIDI mapping views for X axis (left-right) panning of the first and the second faders. They are set in fader bank channel strip class with the fader bank number in order.

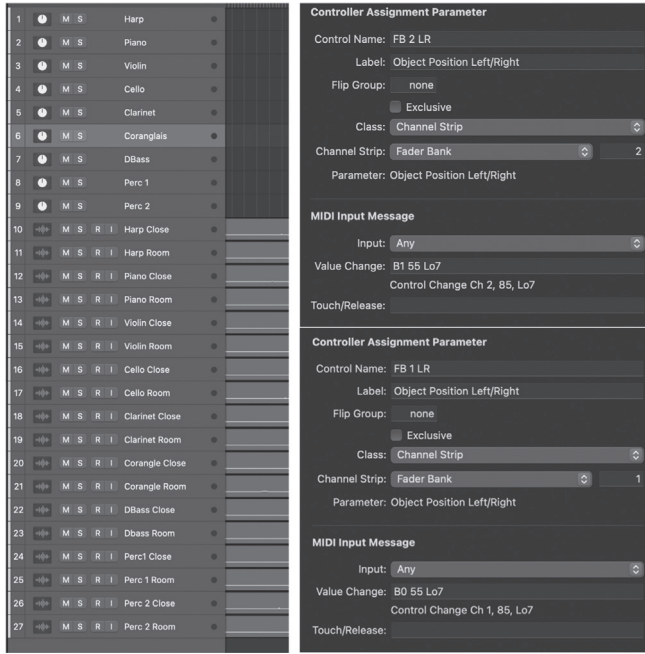


Figure 13. a) Channel order in Logic Pro (left) b) MIDI mapping in Logic Pro (right)

The MIDI input section shows that the controller numbers are the same as in the Max/MSP ChCtrl objects (CC85) and the channel numbers are the same as the channels' fader bank number in Logic Pro. The value change box displays the MIDI controller assignment as a sequence of hexadecimal bytes. In the first channel's assignment's, B0 is MIDI CC Channel 1, 55 is 85. Lo7 indicates that the MIDI signal is 7-bit, and the variable parts of the messages are the lower 7-bits (aka. Least Significant Bit or LSB).

Conclusion

This project has its unique requirements which basic Max/MSP usage is enough to work on. There are other developments planned for future applications. Z axis cannot be controlled in this version of "PannerBank" because the height information is not used for the realistic approach of this project. As it is explained in the previous sections, the main approach is to create an illusion of walking among the objects staying on the ground. Heights of objects does not change in the actual exhibition, so does in the musical imitation. Even though adding the sense of angle change between ear and the source according to the distance was thought at first, that plan was left because of the technical limitation of the technology. The source should sound like it is at a position which is lower than the listener. However, the lowest position of a source in

Dolby Atmos format is the ear level. Dolby Atmos has a semi-sphere 3D sonic environment. It is impossible to locate a source to a position which is lower than the ear level. Nevertheless, since the height of the sources cannot be changed, automation for Z axis was not necessary in this project. It can be adjusted within the DAW if it is needed in this type of aesthetic approach.

Using 2D objects -as “nodes” is- is not comfortable while working on 3D audio, especially with the height information. It is not clear to understand all 3 dimensions in a single view. It is possible to have three views from different angles, which show top, side, and front view. It can depend on a switch to change a view or three different views at the same time in the same window but both options have their own disadvantages. Having three different views could not be space-efficient for the application. If a switch is used to change the view, it is not easy to understand which source is in front or back. The perspective is not easy to achieve in 2D views. To solve these problems, OpenGL support of MaxMSP can be used to create better visuals, which interpret the sense of sonic space in a more intuitive way. The perspective between the objects can be easier to understand. 3D figures rather than just circles can be used to have a more accurate display. Moreover, “jit.phys” objects can be used for different approaches with its advantages of collision detection and body dynamics. The sources can be automated as if sounds fall with gravity or collide with other sources and make them move.

The OSC protocol can be used to connect other programs along with MaxMSP to create visual expressions. It would create new experiences for both musicians and audiences. Using a daily device such as a regular cell phone is one of those developments currently being worked on. The researcher is working on another MaxMSP patch using a smartphone’s sensor data to pan the musical sources in the mix. In a piece named “Pieces and Parts, Part 1” composed by Nihal Saruhanlı, which was recorded, and Dolby Atmos format mixed by researchers of this paper, sounds of flying birds in the mix were panned by moving a cell phone.

The starting point of the project was to show the possibilities of programming to the sound engineers and the sound engineering students who are mostly trained only in music and music production. Although this project focused on panning, the data can be routed to any other musical parameter. Max MSP’s other features such as audio or video processing allow engineers and producers to create their own tools. The data can be connected to the visuals or lighting systems in live productions. Patches can be used to change the music and visuals together and make the experience more immersive for the audience in a concert venue. Hence, details about the connection between the applications are given in the paper to create another technical approach to connection possibilities.

This patch works well enough to fulfill its function in the abovementioned project. However, some other features should be supported. First, it is not possible to see and revise the current places of the objects in this patch because of the feedback

support problem of Logic Pro. There is no option to save and recreate the walking path in the patch. Also, further technical developments are needed to make mixing easier. There should be a transport section on the patch to start/stop and to see the time code coordinated. The size feature of Dolby Atmos can be controlled via another section in the patch. Lastly, the elevation feature cannot be controlled by the current version of the patch. It was not implemented here because there is no elevation processing needed in the project.

Three-dimensional audio production gives many different possibilities to the producers to expand the sonic experience. Main aesthetic approach is placing the listener to the center; however, dynamic listener panning with steady objects is one of the features, which was not possible before. Moreover, modern technologies like head tracking can make experiences interactive even with the standard streaming services without any need for more than a pair of headphones.

References

- Apple Inc. (2023, 05 12). Controller assignments Expert view Value parameters in Logic Pro. Retrieved from Apple Support: <https://support.apple.com/en-gb/guide/logicpro/ctls71c308ee/10.7.5/mac/12.3>
- Barry, R. (1975, February). Four Channel Sound Systems. *The Michigan Technic*, XCIII(4), 6-10.
- Boren, B. (2018). History of 3D Sound. In A. Roginska, & P. Geluso, *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (pp. 40-62). New York: Routledge.
- Tsingos, N. (2018). Object-Based Audio. In A. Roginska, & P. Geluso, *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (pp. 244-275). New York: Routledge.
- Şekerciler, H., & Yalkın, A. (2023, 07 15). Mindflow. Retrieved from ha:ar: <https://www.wearehaar.com/mindflow>