HELLENIC MEDITERRANEAN UNIVERSITY

SCHOOL OF MUSIC AND OPTOACOUSTIC TECHNOLOGIES

DEPARTMENT OF MUSIC TECHNOLOGY AND ACOUSTICS

Postgraduate Thesis

Pix2MIDI: An Original Application for Converting Color Data to MIDI

Megaklis Vassis, MTA37

Supervisor: ***Nikolas Valsamakis***

A Thesis in the Field of Audio Technology for the Degree of Master of Science in Sound and Music Technologies
Hellenic Mediterranean University

Rethymno, July 2024

Acknowledgments

I would like to thank my professor for his invaluable support and feedback during the whole process of writing this thesis. I am also extremely grateful to my partner for her endless patience and emotional support.

Abstract

'Pix2MIDI' is an original standalone application programmed in the Max visual coding environment, which works as a MIDI controller by using image color data and converting it to MIDI events. The main goal of it, is to provide another way of converting pictures to music, similar to electronic hardware and software implementations of recent years. This thesis is not only focused on the analysis and presentation of the application, but also on the examination of electronic contemporary music synthesis of the past 140 years from a historical standpoint, as there have been numerous innovations that incorporate drawn images or pictures as a means of input, to be converted to sound. Also, there will be a thorough analysis of recent and commercially available software applications that produce sound through pictures, as well as an overview of their features and an evaluation of the produced sound result.

Keywords: music programming, graphic music, visual music, Max/MSP

Table of Contents

Table of Figures

Introduction

Electronic sound synthesis from media such as glass disks and drawn sketches has been present since the early 20th century, through innovative electronic musical instruments that used techniques to manipulate light and sound. The inspiring Russian musician and innovator Alexander Scriabin through his progressive ideas and unique synesthetic perspective of sound and colors, influenced many artists and engineers who referenced him in their works. Pix2MIDI is an original computer application created in the Max coding environment that outputs MIDI events to control electronic synthesizers, just as a MIDI controller, by using data deriving from the color characteristics of imported images. It is greatly inspired by the mid-20th century electronic instruments that used optical techniques, as well as the post World War II era experimental graphic notation and contemporary electronic sound composition. This thesis is focused on providing a brief examination of Scriabin's theory and an overview of the most important optical electronic instruments and modern software applications that incorporate techniques to convert pictures to music. Additionally, the creative process and every feature set of the Pix2MIDI standalone application will be presented and analyzed on an artistic and programming level.

The first chapter is aiming to provide a brief history overview of how pictures and sound are related through different art forms and innovations over the past 120 years. From the perceptual phenomenon of synesthesia, which merges the senses of sight and sound, to creating music from sketches or producing paintings from audio waveforms, contemporary artists have embraced and interpreted these unique capabilities in their work. As technology advanced rapidly in the 20th century, numerous prototype electronic devices emerged from scientific experiments in telecommunications, physics, and electronics. These devices were later utilized by avant-garde composers to synthesize soundtracks and effects for movies. This chapter includes a presentation of the most important and influential electronic sound instruments, such as the ANS Synthesizer and the UPIX System to name a few. At the same time, contemporary music composers explored new ways of presenting their compositions through graphic notation, which is considered as another example of interpreting music with shapes, lines and colors. The chapter also provides a brief examination of three contemporary music pieces by György Ligeti, John Cage and Krzysztof Penderecki, that were written in graphic notation and are considered to be defining examples of this remarkable avant-garde period in music composition history.

In the second chapter, there will be a thorough examination of five software applications that take images as an input and convert them to sound by implementing

different techniques. These applications are either digital audio workstations or plugin software. Each one will be examined about their history, feature set, sound output and finally their way of managing data to convert images to sound. The applications to be mentioned are: MIDImage 3, Kaleidoscope, The vOICe, MetaSynth and Virtual ANS.

The third and last chapter is focused on the presentation of the original application called Pix2MIDI. Pix2MIDI is an original standalone application made in the Max environment capable of utilizing data from images, in order to control VST instruments and convert them to sound. This application, although influenced by the computer software examined in the previous chapter, it works differently by converting the image color data of every single pixel to MIDI events. Due to this different approach to image conversion, it can be considered as a MIDI controller application as it can be connected to any virtual synthesizer and control its MIDI input. This chapter includes a complete overview of the idea around the creative process of this application and a full presentation of the user interface as well as every of its features. Finally, there will be some configuration examples and ideas about future improvements and implementations.

1. Sound and pictures in 20th century contemporary music

In the past 120 years, artists from many different fields of arts have come to the realization of the possibility to utilize pictures and music in new forms of art that have eventually evolved alongside the rise of the motion picture industry. From the perceptual phenomenon of synesthesia, which blends sensory stimulation between the senses of seeing and hearing, to an actual music piece derived from sketches or a painting produced by audio waveforms, contemporary artists embraced this unique ability and interpreted different ideas in their art pieces. As technology progressed rapidly in the 20th century, a large number of prototype electronic devices that utilized sound and images arose from scientists experimenting in the fields of telecommunications, physics and electronics, which were later used by avant-garde composers to synthesize soundtracks and effects for movies.

1.1. Music composition through Scriabin's synesthesia

The multi-sensory phenomenon of synesthesia was experienced by a number of notable artists in the early 20th century who cleverly blended the arts of music and painting. Through the works of Alexander Scriabin - who was first to associate sound directly to color via his colored circle of fifths - and his creation of the colored keyboard (clavier à lumières), a large number of Russian inventors were influenced by them and tried to create their own instrument versions that incorporated sounds converted to picture or the opposite. His extreme ideas as well as the synesthetic perception of other artists like Wassily Kandinsky, also influenced art movements and audiovisual techniques in the following decades.

Literally meaning to sense at the same time, synesthesia is "[…] a neurological phenomenon that occurs when a stimulus in one sense modality immediately evokes a sensation in another sense modality" (Campen, 2010). When two or more senses get triggered at the same time, for example if someone hears a specific note or listens to a song and simultaneously a color comes to mind or associates individual letters of the alphabet to colors and shades, then that can be a form of synesthesia. The earliest mentioned medical occurrence was documented by German physician Georg Tobias Ludwig Sachs in 1812 through a medical dissertation he published concerning his own albinism[1], in which he also elaborated on the phenomenon of synesthesia. Although there was still skepticism among the scientific community of researchers if this was actually the first case, an article published in 2009 which provides the first

[1] "A Natural History of Two Albinos, the Author and His Sister" (1812) by Georg Tobias Ludwig Sachs.

English translation of the original Latin text, proves that statement (Jewanski et al., 2009; Simner, 2012).

There are two types of synesthesia: the projective, in which the individual sees colors, forms or shapes when triggered and the associative, in which the individual feels a strong connection between the sense that was triggered and the stimulus. This thesis is focused more on projective synesthesia, which occurs more often and is widely understood, as some of the artists that will be mentioned, carried this unique neurological difference. More specifically, a sub-category of the projective synesthesia called chromesthesia deriving from the Greek "χρώμα" (chroma) meaning color and "αἰσθησις" (aesthesis) meaning sensation, describes the experience of color, shape or movement when sound is heard. It can occur when hearing generally sounds in the environment such as raindrops or the slow closing of a door and it can also be connected to more technical musical terms. For example, imagining a differently colored image when a major or minor scale is played, or associating specific musical tones directly to colors. One of the most important and major personalities that were driven by synesthesia in their art, was the musician Alexander Nikolayevich Scriabin (1872-1915).

Scriabin was a Russian composer, poet, virtuoso pianist, philosopher and inventor born in the late 19th century Romantic era. His early music pieces were greatly influenced by monumental figures of Romanticism such as Frédéric Chopin (1810-1849), while for some he was considered to be the last composer of that period and, at the same time, one of the first contemporary artists. He lived in a period of rapid changes in arts, music and technology and for that reason his influences shifted a lot. In the last years of his career as a composer, described as the third period (1907-1915), Scriabin's compositions started to gradually move away from standard forms and functional tonality. Although not completely atonal like the compositions of his contemporary Arnold Schoenberg (1874-1951) at the time, Scriabin managed to create musical pieces that blended well known practices with contemporary ideas. He appealed to the concept of chromesthesia by associating colors with harmonic tones of a musical scale, while establishing a color system out of the circle of fifths. Also, in his famous work "Prometheus: The Poem of Fire" (1910), which premiered in Moscow in 1911, he featured a custom-made musical instrument resembling a piano with colored keys, called clavier à lumières[2]. After returning to Russia from the United Stated where he performed a series of concerts "he was no longer thinking in terms of music alone" (*Aleksandr Scriabin | Russian Composer, Impressionist, Mysticism | Britannica*, 2024). His last unfinished work was called 'Mysterium': a seven-day long event in which the spectators also participate through music, poetry, dancing, colors

[2] Literally meaning: keyboard with lights

and scents, all united in a multisensory audiovisual experience. Scriabin seemed to be too ahead of his time in the description of Mysterium, which he unfortunately did not manage to finish and left behind only a poem and some sketches for the music.

1.2. Avant-garde music pieces written in graphic notation

Many modern contemporary composers of the late 20th century and especially after 1950, have experimented with numerous different and one-of-a-kind techniques for displaying, writing and producing music. Around the idea of visualizing music, compositions that are mostly electronic have emerged from post-modern artists and are written in the form of graphic notation. These compositions use a different form of displaying music, unlike the standard notation system that any musician could reproduce without much additional context. Graphic notation pieces incorporate sketches with straight and curved lines, various sizes of geometric shapes and written notes that evolve in time and most of the time require explanation as to what they represent musically. Some very important contributors among others are György Ligeti, Olivier Messiaen, John Cage, Krzysztof Penderecki and Steve Reich. The following chapters include a brief examination of three music pieces that are considered to be the definition of graphic notation. These very significant compositions incorporate every characteristic that defines this unique way of presenting music.

1.2.1. Artikulation (1958)

Artikulation is an electronic composition by Hungarian-Austrian contemporary composer György Sándor Ligeti (1923-2006). It is arguably the best known of his works in the electronic medium, as it premiered in the 'Musik der Zeit'[3] concert of the West German Radio (WDR[4]) in 1958 and was used by Karlheinz Stockhausen (1928-2007) as an example in many of his lectures on electronic music. According to the composer it is also "the most thoroughly worked out of his three tape pieces"[5] (Levy, 2006). Ligeti prepared and recorded the piece on magnetic tape with the assistance of Gottfried Michael Koenig and Stockhausen, by whose compositional techniques he was also influenced. The composition is called 'Articulation' because it can be heard as an articulated conversation which develops, through a combination of questions and answers, high and low-pitched voices and sudden interruptions.

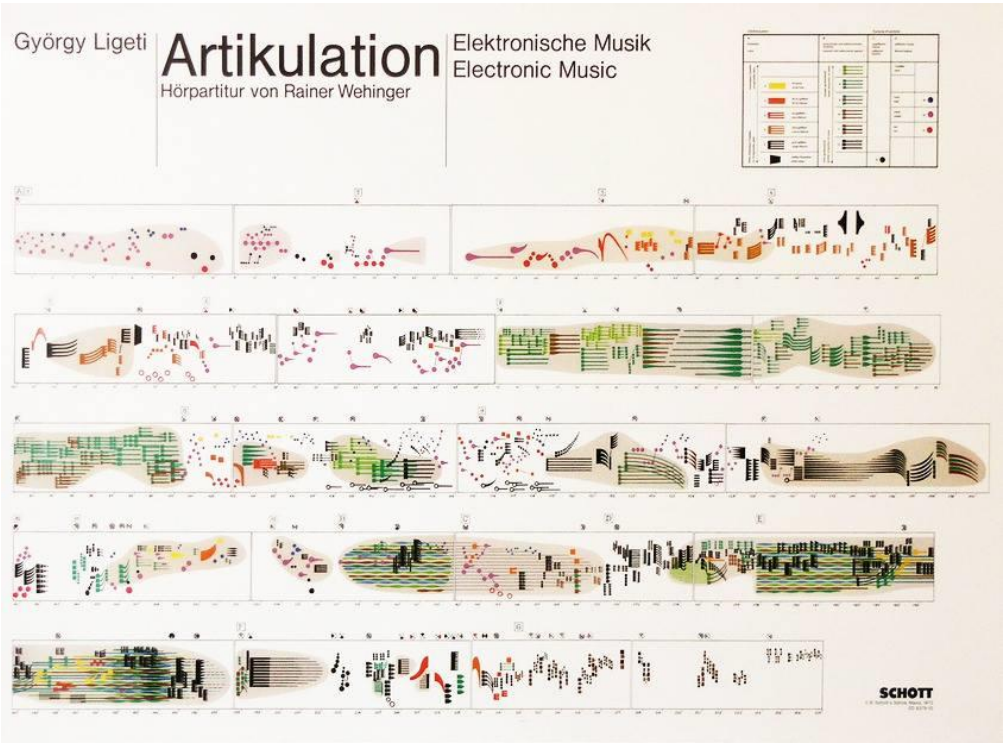
The piece is comprised of different types of sound elements, like sine waves, white noise, impulse generators and filters organized in different series. After pre-recording

[3] Music of Time

[4] Westdeutsche Rundfunk Köln

[5] The other two being Glissandi (1957) and Pièce électronique no. 3 (1957–58)

the sound elements, he categorized them according to sound texture and then grouped the result samples into various categories. The actual samples were then organized into series by a random process. Ligety embraced the aleatoric techniques of Stockhausen by deciding to leave the selection of some sound samples to chance. The original notation of the composition consisted of a large number of printed charts and tables and it was not meant to be read, as it existed only for keeping track of how each sound element was organized. In 1968, graphic designer of State University of Music and Performing Stuttgart, Rainer Wehinger designed the first graphic score of Artikulation (Picture 1). It was an 'aural score', or a score for listening as Wehinger mentioned, which had a distinct and appealing aesthetic to the music piece, was easy to follow while listening to it, and most importantly, was approved by Ligeti himself (Hidalgo et al., 2016).



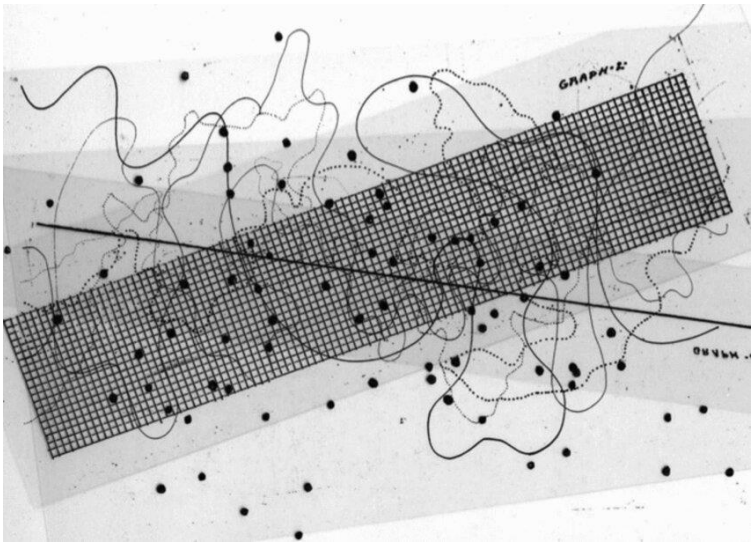
Picture 1: The aural score for Artikulation by Rainer Wehinger (1968)

1.2.2. Fontana Mix (1958)

In 1958, American avant-garde composer and music theorist John Cage (1912-1992), composed Fontana Mix, an experimental electro-acoustic piece. Cage was one of the most influential mid-20th-century composers, as his way of thinking outside the box and implementing unorthodox ideas to his compositions made him stand out of the

rest (*John Cage | American Composer & Avant-Garde Innovator | Britannica*, 2024). Contemporary and experimental music after the 1950s was not necessarily presented to mass audiences via concerts. Instead, most ideas were contained and shared between the inner circles of avant-garde musicologists. Cage was bold enough to host concerts and even present or perform his compositions on television, in front of audiences who were not familiar with such progressive ideas, as he considered any audience reaction, such as chatting and clapping, to be part of the performance.

Fontana Mix premiered in Rome, on January 5th, 1959. Originally called 'Performance Mix', Cage renamed it after his landlady's name while in Milano, Signora Fontana. It is an electronic piece for magnetic tape which consists of 17 minutes of pre-recorded material and can be used in performances of any duration, while a part or parts of the score can also be provided to be used in any instrument combination. The score of this composition is a graphic one, which is composed of ten sheets of paper and 12 transparencies. Each sheet of paper contains drawings of six curved lines in various thicknesses, while ten of the transparencies have randomly distributed points. The last two transparencies contain a grid and a single straight line respectively (Picture 2). By placing one transparency over the other, constructing a combination of curved and straight lines and dots on a grid, "the player creates a structure from which a performance score can be made" (*John Cage Complete Works*, n.d.). Fontana Mix is one of those pieces which is treated more as a manual to the notation for the performer to interpret. A great example of how sketches and geometrical elements can be assigned to musical characteristics such as volume, timbre and pitch, in the same way early electronic sound synthesis instruments worked (Sauerwald, 2024).



Picture 2: Papers and transparencies of Fontana Mix (1958)

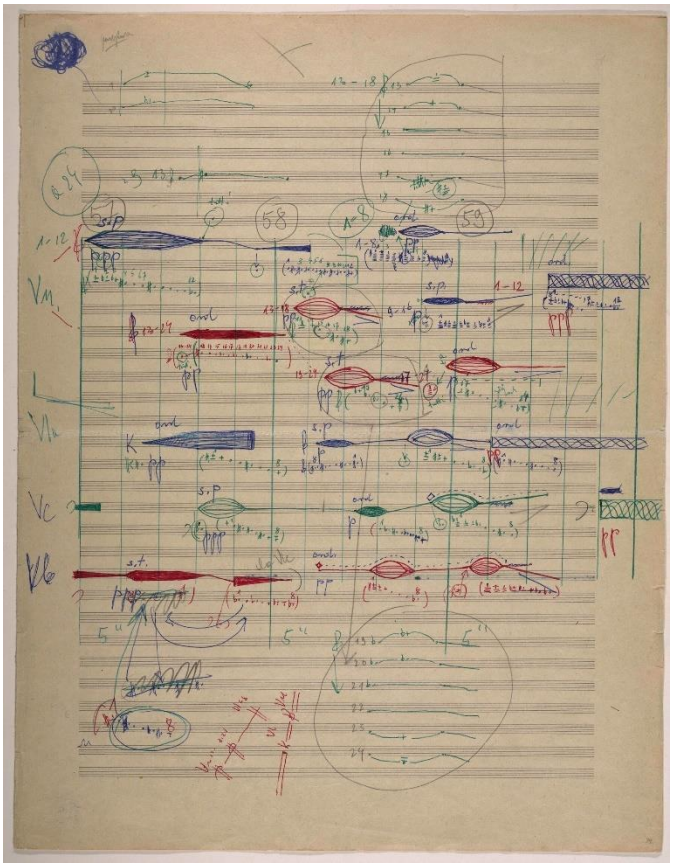
1.2.3. Polymorphia (1961)

Polymorphia[6], is a music piece composed by Polish composer Krzysztof Penderecki (1933-2020) in 1961. Commissioned by the North German Radio Hamburg, the composition was premiered in 1962 by NDR Elbphilharmonie Orchestra[7] and conducted by Andrzej Markowski. The piece is written for 48 string instruments, specifically: 24 violins, eight violas, eight cellos and eight basses. Penderecki was seeking to explore more ways of producing sound by different and unconditional techniques on string instruments after his studying years, thus as the title suggests, Polymorphia emphasizes more on sound timbre than pitch. He focused more on the ability of so many string instruments to generate powerful sound bodies and understood them as timbral materials which take multiple forms and interact with one another.

Composed during his experimental period, the score of this composition is a set of sketches containing lines and thick curves that represent differences in tempo and sound pitch. The written symbols have an abstract meaning; therefore, they mostly represent tone clusters or techniques of producing sound in string instruments such as glissando. "The manuscript (pages) […] are quite impressive visually, both for the aggressive use of color and for the dominating presence of Penderecki's innovative graphic notation" (Monastra, n.d.). Polymorphia accounts for another important example of how musical terms and characteristics can be visualized through symbols on an organized graphic score (Picture 3).

[6] Meaning 'many forms', from the Greek 'πολύ' (poly) and 'μορφή' (morphe)

[7] A radio orchestra employed by Norddeutscher Rundfunk (NDR) German radio and television broadcaster



Picture 3: A manuscript page of Polymorphia (1961)

1.3. Optical sound synthesis

In the early 20th century, some attempts were made to synthesize sound from light by various individuals and researchers from the fields of mathematics, physics, electronics and telecommunications. In the contemporary realms of the arts, film and music, while heavily influenced by the constant technology advancements, several prototype devices were made that used hand drawn animations, electromagnetic waves or light in order to synthesize sound or even the opposite, meaning to paint pictures by generating sounds. Historically, many of the inventions came from scientists of the former Soviet Union and were initially not very well known beyond the borders of their own country, until famous Russian contemporary and avant-garde composers started using them to write music, while their compositions also accompanied many movies. Ultimately, these technologies were mainly developed for recording soundtracks as each and every one of the innovators provided music scores to the film industry using their unique graphical sound techniques. It is clear that the connection of optical sound technologies with filmmaking - especially the Visual

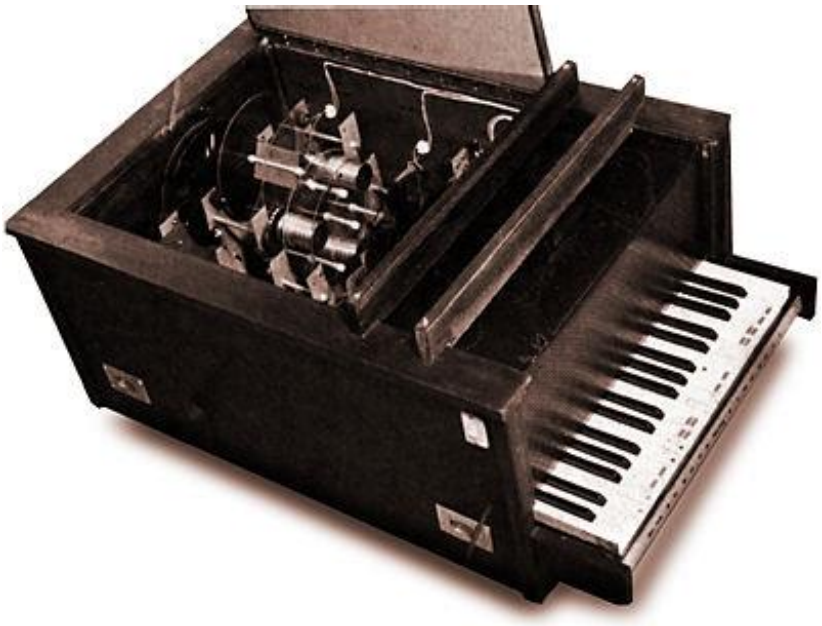
Music movement - is very close in the mid-20th century. From the 1920s and 1930s through the works of director William Norman McLaren, to the late 1960s and 1970s filmmaking and animations made by John Whitney, Len Lye, Oscar Fischinger and many more, the relationship between cinematic techniques and direct optical sound synthesis was more profound. In the following chapters, some of the most important and influential inventions involving sound and pictures will be analyzed in ascending chronological order.

1.3.1. Optophonic Piano (1916)

Vladimir Baranov-Rossiné (1888-1944) was a Russian-Ukrainian avant-garde painter, sculptor and inventor. In 1916 Rossiné started working on the Optophonic Piano, a one-off electronic optical instrument that he presented at various exhibitions of his own paintings (*The 'Optophonic Piano', Vladimir Rossiné, Russia and France. 1916 – 120 Years of Electronic Music*, n.d.). This custom-made instrument was used in two concerts he performed with his wife at the Meyerhold and Bolshoi theatres in 1924, as well as many more in Paris in which he emigrated after he left the Soviet Union in 1925. The Optophonic Piano operated by directing a bright light source through a series of revolving colored glass disks - painted by Rossiné himself - filters, lenses and mirrors, and whilst using the filtered information, it produced sounds and projected revolving patterns onto a flat surface like a wall or ceiling (Picture 4).

Picture 4: Painted glass disk of the Optophonic Piano

The piano keyboard was used to control the order and combinations of the various disks and filters. While the light was getting filtered through the painted disks, a photo-electric cell picked up the variations in opacity in order to control a single oscillator's pitch. By utilizing only one oscillator, the instrument produced a continuous tone that varied in pitch, alongside the rotating circular projections which simulated depth, resembling a kaleidoscope (*Optophonic Piano - Monoskop*, n.d.). With sound and picture happening at the same time, the Optophonic Piano was more of an interactive audiovisual art installation than just an electronic musical instrument (Picture 5).



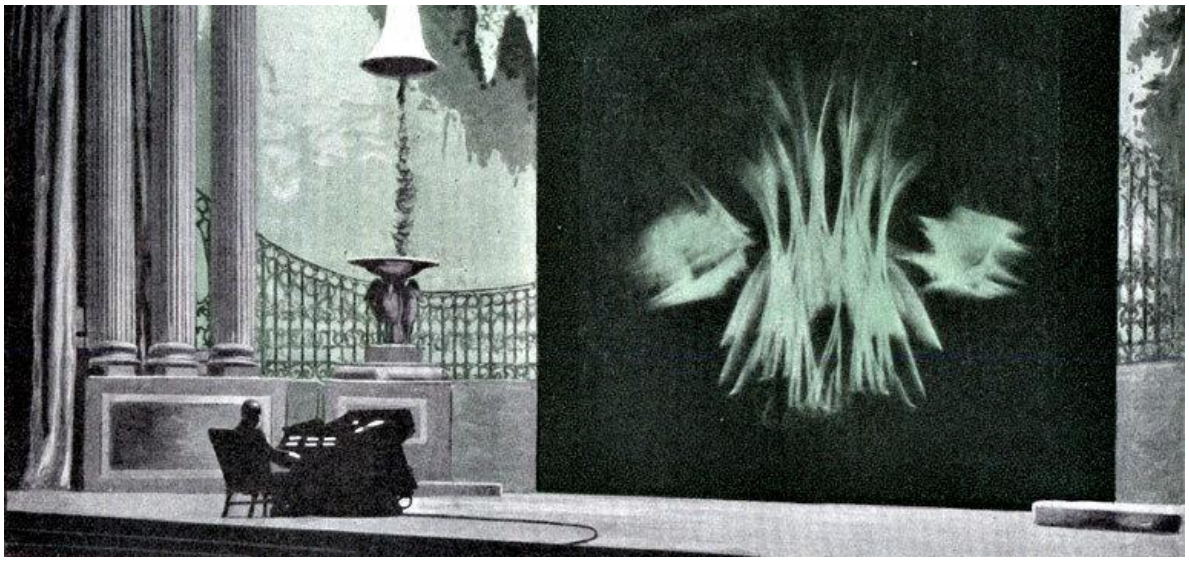
Picture 5: The Optophonic Piano (1916)

1.3.2. Clavilux (1919)

Thomas Wilfred (1889-1968) was a Danish-born visual artist, inventor, designer and musician who created the Clavilux color instrument and is connected with the term 'Lumia' which was his term for explaining the form of art that incorporates light patterns. The Clavilux, meaning 'light played by keys' deriving from the Latin words 'clavis' for key and 'lux' for light, was a mechanical device that used hand-painted color patterns on rotating glass disks. Wilfred began building the first version of Clavilux, named Model A, from March to May 1919, while in the following years he revised it by making many more alternate versions and sixteen smaller home models which he called 'Clavilux Juniors'. The reason behind the smaller versions is that he imagined that the abstract Lumia art created by the instrument will accompany

music concerts and movies in the near future. After moving to New York and establishing an Art Institute of Light, he toured around the world performing Lumia concerts in the United States and Europe, notably at the Art Déco exhibition in Paris (*The Dream of Color Music, And Machines That Made It Possible*, n.d.). To this day only a few Clavilux Juniors exist in private collections, but also two full-sized ones, the Model E (1924) and Model G (1936) were rescued in 2003 in order to get fully restored to working condition.

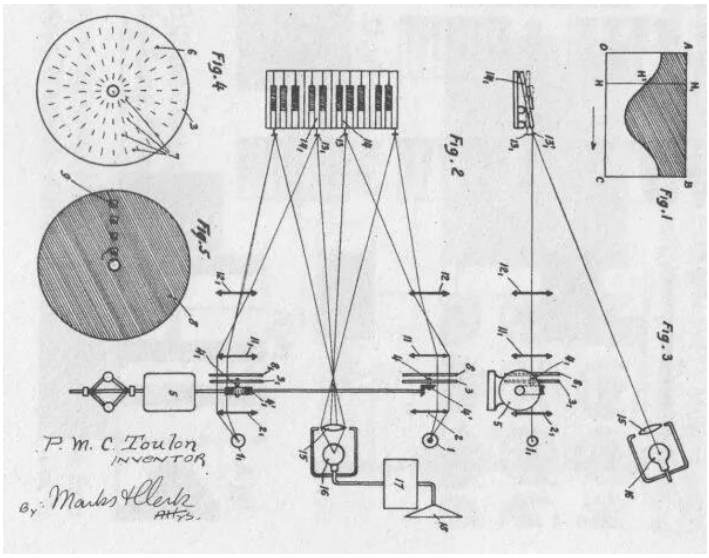
Wilfred's Lumia work is used to describe an emerging art, a composition of light and color which changes slowly with time and is organized around these three basic factors: form, color and motion. The actual mechanism of the Clavilux is surprisingly sophisticated for the time it was invented. It used a strong neutral white beam of light through an arrangement of lenses and tinted screens and then projected it onto a wall. Depending on the variant, the instrument used three units that correspond to manuals on a pipe organ. Each of the units contains a set of sliding keys divided into three groups: form keys, color keys and motion keys. Each key can be set to one of 100 positions and one could use its own notation book to play the instrument (Collopy, 2000). In place of a note head the keys would have chords with numbers and lines on a staff as on a piano. The light which gets intercepted by the lenses is built into a form by the form keys. The projections that get constructed are moved rhythmically by adjusting the motion keys and finally the independent colors or color combinations are introduced by the color keys. Although the Clavilux is more of another machine which represents the Visual Music tradition rather than a sound-producing standalone instrument, the graphical sound techniques and the technologies it implements are significant (Picture 6).



Picture 6: An illustration of Thomas Wilfred playing the Clavilux

1.3.3. Cellulophone (1927)

The Cellule Photo-électrique or Cellulophone made its debut as a prototype in France in 1927. Invented by the French engineer Pierre Marie Gabriel Toulon and aided by the electronic engineer Krugg Bass, the Cellulophone was an electro-optical tone generating instrument which resembled an electronic organ. The instrument's sound was created using a photo-electrical technique by projecting a light beam through equally distanced slits on a disk that rotated in variable speeds. They used four tungsten filled lamps to generate constant light beams which then were picked up by a selenium meter's[8] photoelectric cell after getting directed via mirrors and lenses through the slits of the rotating disk. Using this method, the machine generated an audible voltage pulse from a single vacuum tube oscillator. Therefore, the note frequency of the output signal was determined by the speed and number of slits on the rotating disk, with 54 slits being the lowest note, while differently shaped disk filters designed by Toulon were added to modulate the timbre of the instrument. Much like the majority of these inventions, it also used a keyboard to control pitch. The Cellulophone had many innovations around the idea of producing sound from a light source, but despite that, its use of a single rotating disk to generate all twelve notes in an octave made the instrument out of tune every time (Darter & Armbruster, 1984) (Picture 7).



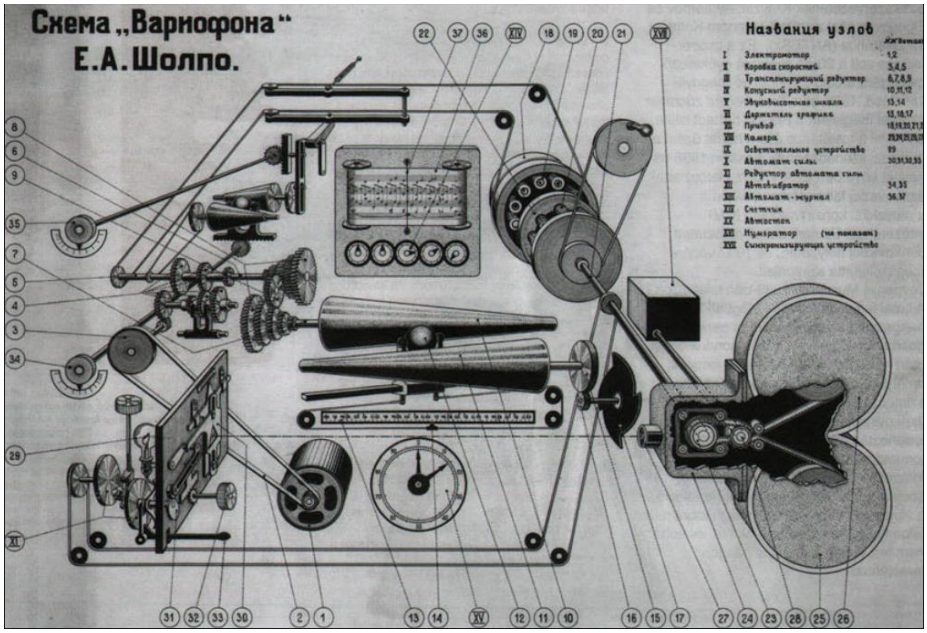
Picture 7: Pierre Toulon's original 1928 patent of the Cellulophone

[8] A photoelectric light meter, used to measure light exposure

In his career Pierre Toulon held a large number of patents related to inventions around circuitry, communications and sound reproduction for example, electrical relays, large screen televisions and audio amplifiers, while in 1930 he proposed a technique of synthesizing speech by using optical film and a rotating drum. Toulon's Cellulophone also won the second prize in a competition held by the French National Office for Scientific and Industrial Research and Inventions in 1933 (*Recherches et Inventions | 1934-10-01 | Gallica*, n.d.).

1.3.4. Variophone (1930)

The Variophone was a photo-electrical electronic instrument developed by Evgeny Alexandrovitch Sholpo (1891-1951) together with the composer Georgy Rimsky-Korsakov (grandson of Russian composer Nikolai Rimsky-Korsakov) in 1930 at Lenfilm Studio Productions in the city of Leningrad, located in the then Soviet Union. The device was the longest living project of his own laboratory until his death in 1951 and a proud achievement, as he was experimenting with graphical sound techniques for many years. The Variophone was an optical synthesizer which was based on the method of photo-optic sound recording used in cinematography in order to construct visible images of sound waves or even the complete opposite, meaning to synthesize sounds from sound waves drawn on paper (Picture 8). The first practical result was achieved one year after the initial experimentations in 1931, aided by Rimsky-Korsakov (1901-1965), with the device being used to produce soundtrack for the film 'The Year 1905 in Bourgeoisie Satire' and again in 1932 for 'A Symphony of Peace' and 'Torreodor'. In the following decades Sholpo produced many sound films and artificial soundtracks for both films and cartoons using the Variophone. At the siege of Leningrad in 1941, the instrument was destroyed, while after World War II Sholpo became the director of the new 'Scientific-Research Laboratory for Graphical Sound' with Boris Yankovsky at the Leningrad's State Research Institute for Sound Recording. After being removed from this position and moving his laboratory in Moscow, he worked on the fourth and final version of the Variophone, experimenting with musical intonation and the temporal characteristics of live music performance, but he never finished it as he passed away in 1951 after suffering a long illness and the laboratory was permanently closed.



Picture 8: Diagram of the Variophone by Evgeny Sholpo

Being a photo-electric audio synthesizer, the Variophone used sound waves cut into cardboard discs which were rotated alongside a moving 35mm movie filmstrip. By filming again these two combined, the movie could be played back with the produced sound on a normal movie projector. In that manner multiple layered tones and sounds could be added with the method of overdubbing. Evgeny Sholpo, after his research and experimentation around performer-less music, designed the instrument with particular methods "to allow the composition of lengthy polyphonic pieces of music" ("The 'Variophone' Yevgeny Sholpo. Russia, 1932," 2013). The polyphonic capabilities of this device were at first limited to six voices in the early versions by combining monophonic parts, while by the late 1930s and 1940s the number doubled to up to twelve voices recorded on the soundtrack area of the filmstrip. The operation of the Variophone as an instrument that represents the graphical music movement, was also influenced by the works of avant-garde composer Arseny Avraamov (1884-1944) who created the first hand-drawn soundtracks in 1930.

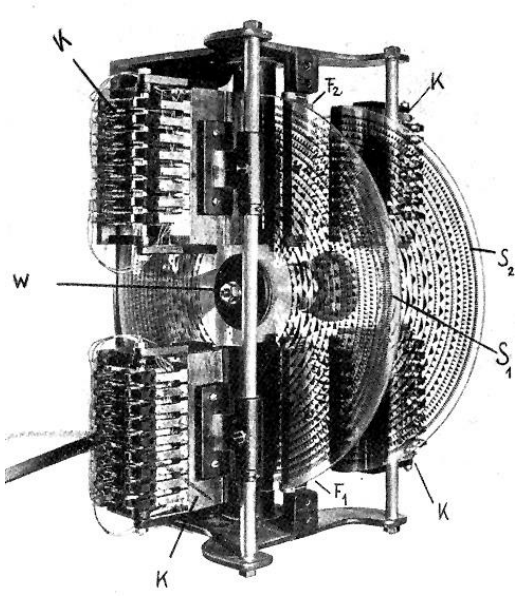
1.3.5. Welte Light-Tone Organ (1936)

In 1936 Edwin Welte (1876-1958) of the famous German mechanical musical instrument and organ manufacturer Welte-Mignon, designed the 'Welte Licht-Ton Orgel' (Welte Light-Tone Organ), an electronic instrument that used electro-optically controlled tone generators. It produced sound by filtering a light beam through glass disks with printed patterns. Welte began experimenting with different materials on

prototype optical disks as early as 1925 before finishing the production version of Lichttonorgel that used glass ones. On the 6th of November 1936, the organ premiered at the Berlin philharmonic by Kurt Grosse, a virtuoso German organist and was covered by the media and received a large number of positive reviews, notably by the Völkische Beobachter newspaper that mentioned "A unique miracle, perfection in the entire realm of concert instruments" (9. November 1936) (Picture 9). Soon after the well-known Telefunken company negotiated a production deal but, in the end, it did not come through after it was discovered that Welt was married to a German Jew. Although after the war he continued trying to release the instrument commercially, he stumbled upon the increasing competition from systems that were cheaper and more efficient and finally completed only three production models of the Lichttonorgel.

Picture 9: The "Welte Licht-Ton Orgel" (Light-Tone organ) (1936)

In order to produce sound, the instrument used 12 glass disks, much like on the Variophone, with each one having 18 different waveforms printed on them in concentric rings (*Lichtton Orgel (1935) | Digital Music Academy*, n.d.). A light beam was filtered after passing through these tone disks and then captured by a series of photoelectric cells for controlling the sound timbre and pitch. In the same way the standard organ utilizes different timbres, the Welte Light-Tone Organ produced three different timbres for each note of the keyboard across all octave registers (Picture 10).



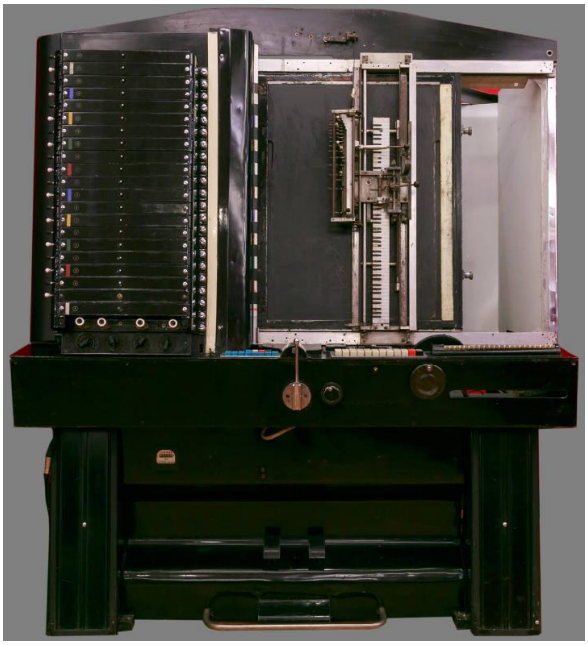
Picture 10: Glass discs of the Light Tone Organ with printed waveforms

1.3.6. ANS Synthesizer (1958)

Russian optical engineer Yevgeny Alexandrovich Murzin (1914–1970) invented and developed the ANS synthesizer, a photo-electronic instrument, over a span of twenty years between 1937 and 1957. In 1938 he designed a system that synthesized complex musical sounds by imputing limited numbers of pure tones, which he proposed to composers in order to perform music without the need for musical instruments or musicians. Inspired by the film industry at the time, this invention was based on the method of photo-optic sound recording used in cinematography, which was capable of constructing visible images from sound waves as well as synthesize sound from a drawn spectrogram. Although the idea and the principles around it were simple to understand, it took Murzin around twenty years to finally assemble a working machine as his engineering work was unrelated to music and he approached the development of the ANS synthesizer as a hobby. In 1958 he gathered a group of engineers and musicians and opened a laboratory, so his idea and design came to life (Kreichi, 1995).

The ANS synthesizer took its name from the initials of the famous Russian composer Alexander Nikolayevich Scriabin, honoring his color-sound theories and influence in the graphical sound movement of avant-garde composers and theoreticians. The instrument required a photo-optic generator, an original process that Murzin developed, which utilized generated sine waves that were printed onto five rotating glass discs. Each disk contained 144 individual very small graphic representations of

sound waves that were drawn by hand, called optic phonograms. These represented pure tones or sound tracks, while the track located the closest to the center of the disk produced the lowest frequency and the track farthest to the edge produced the highest. By projecting a light beam through the rotating disks onto a photovoltaic cell which resulted in a voltage tone corresponding to the frequency drawn on the disk. With these five disks rotating at different speeds, the ANS synthesizer was able to produce 720 discrete pitches (microtones), spanning ten octaves and covering the whole audible spectrum (Picture 11).



Picture 11: The ANS Synthesizer

The user interface of the instrument consisted of a glass plate covered by a layer of black mastic resin which worked as a drawing surface. As the user marked the surface by scratching the black layer, light was allowed to pass through and then be captured by the photocells and get transformed into sound (Picture 12). In other words, it generates sound depending on what is drawn. This non-drying mastic layer allowed for instant editing and correcting of the resulting sound so the users can smear off unwanted sounds or add missing ones. A bank of twenty photocells sits in front of the glass plate which send signals to twenty band-pass amplifiers, each with their own gain adjustment switch. This plate works like a spectrogram where the vertical axis corresponds to pitch and the horizontal to time, and can be moved manually to adjust the scanning speed even down to zero, altering the duration of the composition. The ANS synthesizer is completely polyphonic and capable of generating all 720 of its pitches at once if needed.

Picture 12: The plates used in the ANS Synthesizer

Many contemporary composers of the 20th century used the ANS in their works as well as for movie scores. The Russian composer Edward Artemiev used the ANS synthesizer to write abstract sci-fi scores for Andrei Tarkovsky's movie Solaris (1972) (Matthew, 2021). Some of the compositions that were recorded and released among others were "The Stream" by Alfred Schnittke who experimented with the machine from 1968 to 1970, Edison Denisov's "The Singing of the Birds" and Sophia Gubajdullina's "Vivente-nonvivente". The first vinyl LP that featured recordings of the composers mentioned above was the "Musical Offering" released by the Russian label Melodiya (C60 30721 000) in 1990. A while later, in 1999, Electroshock Records, another Russian label released a series of electro-acoustic sampler records with the fourth issue called "ELECTROACOUSTIC MUSIC VOL. IV. ARCHIVE TAPES. SYNTHESIZER ANS. 1964-1971" (ELCD 011), being devoted to archive recordings of the synthesizer, as well as two pieces by the ANS expert and composer Stanislav Kreitchi (1936-2021) ("The ANS Synthesizer Pt2.0 ANS Discography," 2007).

The ANS synthesizer is a very important and monumental device related to graphical sound synthesis. Alongside similar soviet era engineering marvels connected to electronic sound composition, it sparked the interest of many contemporary artists, mainly musicians, and was highly praised for its unique design and sound output.

1.3.7. UPIC[9] system (1977)

Famous avant-garde composer, architect and engineer Iannis Xenakis (1922-2001) developed the UPIC system in 1977 with help from the computer engineer Patrick Saint-Jean at the Centre d'Etudes de Mathématique et Automatique Musicales (CEMAMu) in Paris. The UPIC was a computer-based controller for digital music based on early Russian optical synthesis techniques and equivalent analogue instruments for graphical sound synthesis such as Yevgeny Murzin's ANS Synthesizer. Xenakis saw the artistic potential of computer systems as early as in 1961, when he used a program made to calculate mathematical probabilities in order to generate algorithmic scores for his famous piece 'Metastaseis' on an IBM computer. As technology progressed rapidly in the late 1960s and with more powerful hardware available, he experimented on ideas that ultimately led to the realization of the UPIC system. Xenakis imagined a composer who is unrestricted by complex software and conventional music notation, opening up the possibilities for people with no formal musical training to create music in a more direct way. In order to prove how the system could work educationally and to demonstrate the blending of music and architecture as a creative practice, he often invited young people and non-experts to see and play the UPIC within educational activities that were organized by CEMAMu (Xenakis et al., 2020).

The UPIC system used a large high-resolution digitizing tablet as an input device where the user could draw upon and every action was then displayed on a CRT screen. A Solar 16-40 mini-computer analyzed the input data and generated digital sound (Picture 13). Although at first the sound synthesis process did not occur simultaneously to the data input, later versions of the system were developed for real-time composition making the system more interactive and utilized data in the form of digitally sampled waveforms instead of synthesized tones (Crab, 2014). The UPIC system can be thought of as the early digital version of the ANS Synthesizer as the composer interacts by drawing on a X/Y axis to generate and organize sounds.

[9] UPIC stands for Unité Polyagogique Informatique du CEMAMu

Picture 13: The UPIC System hardware

As of the later versions, wavetable samples which were created by drawing straight or curved lines, were loaded into a memory buffer and then the program started picking up data by looping the wavetable continuously. This sequence of samples formed a digital audio signal which was finally sent to loudspeakers after passing through a digital-to-analog converter (DAC). The unique nature of the system lies in the interpretation of said drawn lines. "[…] once a line or curve has been drawn, nothing is really decided about its interpretation as a signal: it can be used as a wavetable in the audio level process, as previously mentioned, but it can also be used as a control signal. In the former case it affects the timbre or sound color, in the latter, it works as a low frequency oscillator (LFO) - driving the amplitude (envelope shaping) or the frequency (fixed pitch or glissando) in the sound being synthesized" (Di Scipio, 2023). This method of interpreting data requires the composer to check many times if the drawings actually make musical sense. The connection between what is drawn and the sound outcome is not always predictable, so in that manner, most of the time the drawing needs to be adjusted first. Besides, this stochastic process is a characteristic of Xenakis' compositional practices.

The UPIC system has been used by many composers in their creative works, including Xenakis himself with his first piece being Mycenae Alpha (1978), a part of the multimedia show Polytope de Mycénes (1978), which he composed completely out of computer-generated sounds made on the system. Other pieces include Saxatile (1992) by Jean-Claude Risset and Illusions in Desolate Fields (1994) by Takehito Shimazu. In 2005, a 2-CD compilation of works composed with the UPIC at CCMIX[10] titled "Xenakis, UPIC, Continuum" was released by Mode Records of New York, a label focused on contemporary and avant-garde music. It contained nine recordings

[10] An abbreviation for Center for the Composition of Music Iannis Xenakis (2000), formerly Les Ateliers UPIC, CEMAMu (1972)

of compositions by artists such as Curtis Roades, Shimazu, Risset, Xenakis and more, with some of the compositions being released for the first time (*CCMIX Paris*, n.d.; Mattis, 2002).

2. Image to sound software

This chapter is focused on historically important and currently available digital audio workstations and VST plugins that use an image as data input in order to convert and export it to audio. These applications vary from simple dedicated software to complex and full-fledged audio editors and each one utilizes different techniques of managing image data to their advantage. In the following sections there is an overview of their history, feature set and capabilities.

2.1. MIDImage 3 by Brute Force Programing

Pioneering in the field of synthesizing sound and music from image data input, MIDImage is an innovative computer software capable of manipulating MIDI data controlled by pictures. As an application, it provides maybe the most direct connection of pictures to music as it includes a variety of settings regarding musical aspects while being easy to understand and create sounds with. Although it never flourished commercially and was left behind recent updates and user interface improvements, currently standing at version 3, MIDImage is a great experimental entry level creation tool.

2.1.1. History

In the late 1990s, Daniel Wojcik got involved with MIDI synthesizers for music creation and quickly became familiar with computer applications of that time such as Paul Whalley's QuaziFractal Composer and Lars Kindermann's MusiNum – The Music in the Numbers, both released in 1999 (Wojcik, 2023). The aforementioned applications were centered on using mathematics as a means for composing music and soundscapes and gave him the idea of mapping numerical values of MIDI note data ranging from 0 to 127 to the values extracted after analyzing colors from the pixels of an image, ranging from 0 to 255. This idea remained to be seen whether it could lead to a valid musical result or just sound chaos. Drawing from his experience, Wojcik started working on this idea in 1999 releasing his first software named 'The Loxound Musical Pixelator' which later became known as MIDImage, a standalone application aiming to produce sound from image input. At the time of writing, this application has evolved with adding many features coming from standard music editing software and is available to download from the official website as version 3.5.4.0 (2013). Although being officially a demo, this version comes fully loaded with features and is free to use. MIDImage is more of a niche music creation software than a generic sound plugin, musical examples of which can be found throughout music distribution platforms from Wojcik himself under the alias TDIS (The Door Into Summer).

2.1.2. Overview

The way that MIDImage works is by reading an image and analyzing the colors of each individual pixel according to the RGB color space, as every pixel has a value for the three colors red, green and blue ranging from 0 to 255. The values then get translated to MIDI events in order to convert the image to sound and because the MIDI protocol can handle pitch, volume, velocity, channel and a variety of other sound characteristics, the end result can be very interesting musically. The application was differentiated from other similar software because of the capability to select a musical scale for the end composition with many options such as the basic major and minor scale and many musical modes, such as Dorian and Phrygian to name a few. Wojcik achieved this by creating a series of "correct" notes for every scale or mode so that by comparing them with the whole image output data, only pitches of the selected scale will output sound while the "wrong" ones will be rejected or translated to music rests. The graphical interface is very user friendly without any hidden menus, giving the user an easy-to-use environment to work on. There are three main editing modules regarding image, scan and music settings.

- Image

The user can load any image which automatically adjusts to the resolution of 200x200 pixels for uniformity purposes. Additionally, there are some options for basic image editing such as flip horizontally or vertically, rotate right or left and shrink.

- Scan

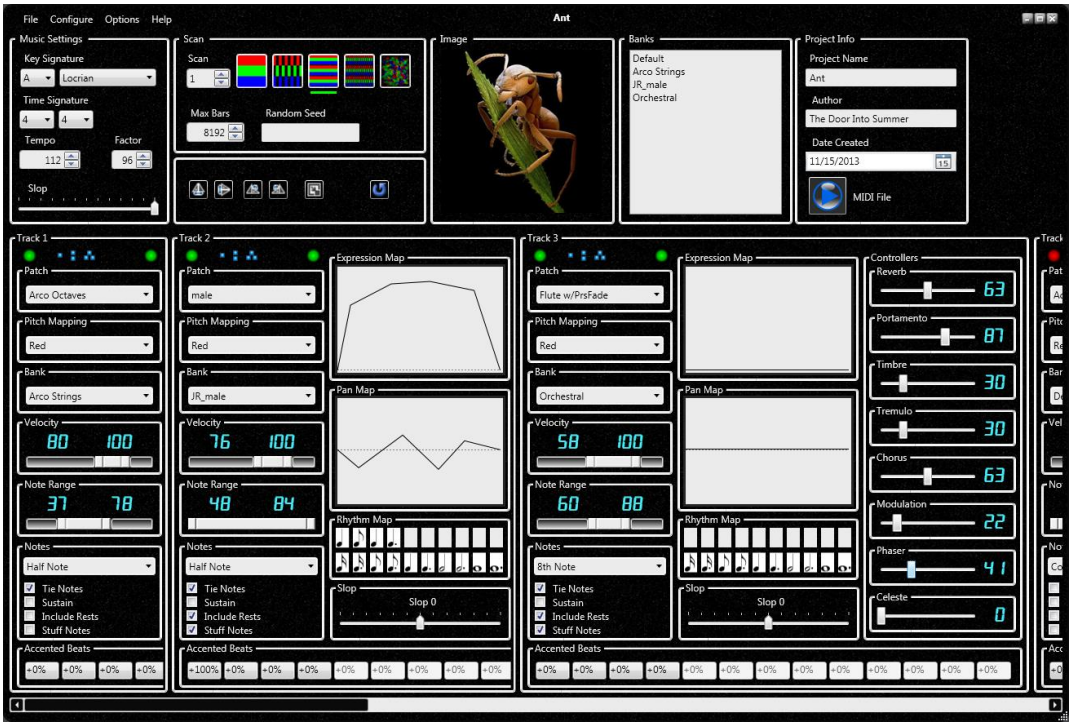
A very important aspect that affects the end result relies on which way the image will be "read" or scanned. There are five options: linear, averaged linear, banded, average banded and random with an additional seed value, while the step parameter defines how many lines of pixels will be taken into consideration on the averaged linear and averaged banded modes. Also, in this section the length of the composition can be altered by inputting the max number of bars, although the exact duration will depend on tempo and time signature, as well.

- Music Settings

Respectively important are the music parameters which will add character to the composition. There are options for key signature as well as music modes, time signature and tempo.

The above three sections form the basic structure of the sound result while all the other available parameters focus more on quality characteristics like expression,

velocity and panning. By default, there are sixteen individual tracks ready to play, although only three of them are initially active (Picture 14).



Picture 14: MIDImage 3 user interface

2.1.3. Sound Output

In its effort to approach musicality in the compositions, MIDImage restricts itself by methods used for sound creation regardless of the variety of its features. Apart from the tonal and expression characteristics that can be configured in each track, there are also settings for notes and rhythm which alter note duration and rhythmic patterns. The fact that these settings exist per track, especially the Rhythmic Map, means that a different rhythmic pattern can be configured for every single one of them. In an example using three separate tracks with each getting data from different colors amongst red, green and blue, it results in three different melodies created which move in their own rhythmic patterns without uniformity. Each melody can exist on its own, musically following the selected scale, but it just sounds like different melodies from different pieces are played together. There could be significant improvement if there was a global setting for rhythmic patterns that changes after a set number of times, as this can give the impression of minimalism through repetition.

Moreover, it is especially difficult to interpret beat in the composition. Either you select a 3/4 time signature or a 4/4, or 5/4, it does not seem to make a difference. The way to make a time signature or a rhythmic pattern stand out is by accenting certain notes, meaning that some of them are played in higher volume than others to emphasize the start of each measure. There is a way to alter note volume through the Accented Beats setting by percentage, although it creates confusion while the Velocity slider is configured, which is altered randomly. In the end, with some effort we can configure a profound time signature if some settings are disabled or bypassed, but it is not easily noticeable especially to an inexperienced user.

2.1.4. Conclusions

In its core, MIDImage is a supplementary tool for digital signal processing in the armory of a music producer, with its use being special and for specific applications. It is one of those applications which either is actually useful and loved or completely unnecessary and hated by the user (Nechita, 2018). To its credit, it is one of the first modern computer programs that initiated getting sound output from image data input. However, the late updates and the fact that it was programmed by a single individual beginning as a simple project without extending it commercially, made it obsolete in modern features as well as in appearance.

2.2. Kaleidoscope by 2CAudio

Kaleidoscope is an intricate and complex VST plugin made for shaping and manipulating audio samples while at the same time being a tool capable of composing original sounds using images as a way to receive data. Its revolutionary method to also display multiple expression envelopes in two dimensions as a picture as well as the use of resonators made it stand out among the rest, more traditional audio tools. The sound output is very unique and seems to come out of this world, making it an important and powerful plugin for any audio producer.

2.2.1. History

2CAudio's Kaleidoscope is a plugin signal processing tool which works as a sound generator, meaning that you can create completely original sounds without having any MIDI or audio input and also as an effects processor in order to edit pre-recorded sounds. Its capabilities are endless, having a huge library of resonators and room reverberation emulators which can be tuned with mathematical accuracy, while in the sound synthesis aspect the user can import images and control sound after analyzing them. In the hands of the audio producer, Kaleidoscope may seem like a game because in little time and without much experimentation it can lead to interesting and otherworldly sound results, but an experienced user who has

complete understanding of the intricate workings to manipulate sound at will, can add a unique touch to every composition. The first version was released in 2014 and as at the time of writing, version 1.1.0 is commercially available via 2CAudio as a VST plugin for Windows and Mac OS. System requirements are average, while it is optimized to take advantage of multicore computer processors.

2.2.2. General idea

As music is an organized sound, it is based on two main axes: time and frequency. Time in music is interpreted by the tempo, note duration and time signature while the frequency axis consists of pitch values, music scales and harmony. Kaleidoscope aims to rewrite music theory for the information age by giving the user complete control over these sound principles in order to expand on understanding them and explore ways of how rhythm and frequency converge, with one taking the role of the other. Additionally, it offers dynamic control over the sound space at each point in time and frequency, adding a third axis of creative organization of music, which it refers to as “spatial choreography”.

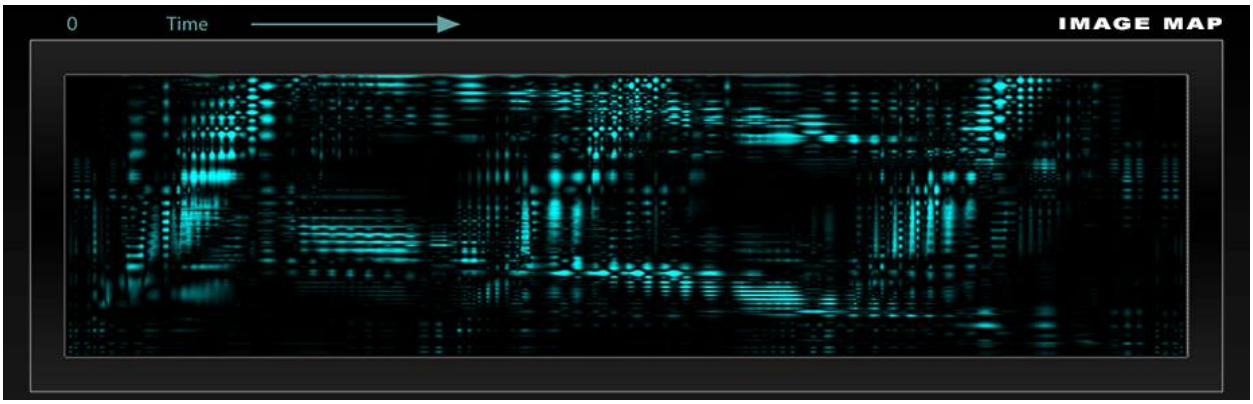
2.2.3. Overview

The main principle of Kaleidoscope is based on resonance. In music, when the term resonance is mentioned, it refers to the way notes resonate with the wooden top of the guitar or the piano, for instance, which leads to these notes being louder and standing out. In practice, if every note would be played simultaneously on an instrument, after a few moments the outcome would not be noise, but certain resonating notes would be more profoundly heard. This effect resembles a prism and how when the white light enters and is filtered, eventually exiting through the other side, the visible color spectrum can be seen. Meanwhile, every acoustic instrument naturally uses sound generating elements such as strings - to be bowed (violin) or plucked (guitar) - or membranes (timpani) and additionally due to the materials they are made of, wood or brass, they resonate in specific frequencies which are called harmonics. The acoustic resonance of a room can be recorded, emulated and applied to a signal, so that the resulting sound has the characteristics of that specific room. In that same manner, the acoustic resonance of an instrument can also be recorded and replicated with mathematical accuracy. Kaleidoscope uses modeled resonators which operate to produce sound or apply higher amplitude to select frequencies. In that setting, the resonator works like a filter which lets these frequencies pass and mutes the others. Up to 512 resonators can be independently tuned and modulated.

A unique feature which makes it stand out is the use of pictures to control sound. Drawing from the visual representation of the spectrum of frequencies in time (spectrogram) and the piano roll used in standard audio editors, it scans the image

from left to right with the vertical axis representing the frequency domain and the horizontal representing time. The luminance of each individual pixel, as well as its position in time and frequency domain, is used to control sound volume. If up to two images are imported using it as an effects processor, numerous different versions of the same sound sample can be created. Also, sound can be produced directly from the image, based on pixel characteristics such as color and hue, essentially converting image to sound. As mentioned, "a single Kaleidoscope instance […] can provide over two million points of stereo automation data to create incredibly intricate performances […]" (*2CAudio - Kaleidoscope | It's A Trip.*, n.d.). Each pixel represents a single sound element ranging in duration - from a minute to an hour to a day or even more - while tempo is manipulated by the speed of image scanning and, given the correct setting, there is the possibility of complicated polyrhythmic patterns to appear.

One of the most important aspects is the Image Map, a two-dimensional plane capable of containing automations and envelopes in the form of a picture. Each modulation curve has 1024 alteration points for better accuracy, but they occupy too much space so no more than one can be represented at once. The software innovates by tilting the modulation curve 90 degrees in the Z axis so that the observer is looking at it from above, essentially becoming a line. That way, multiple lines of independent modulation envelopes are rendered at the same time in two dimensions (Picture 15).



Picture 15: An image map with 256 independent 1024-point modulation envelopes

Differences in luminance of each point in the Image Map control sound amplitude while colors also play a major role. Using the RGB color space the system indicates stereo panning. Different amounts of color combinations create variable hues having two of the primary colors as left and right channel, with their combination being the center. For example, if red represents the left audio channel and green represents the right, the combination of them, being yellow, is the center. This way the user can

visualize stereo panning with ease. Kaleidoscope by default uses the combination of green for the left and blue for the right audio channel, which results in cyan.

2.2.4. Conclusions

There are endless VST plugins for sound sampling and producing but Kaleidoscope, having features as the resonators and Image Map, is one of a kind. Every producer can find their way of using this tool to their advantage, composing bizarre electronic sounds, while the fact that it can make sounds using pictures as direct data input render the possibilities endless.

2.3. The vOICe by Peter B. L. Meijer

The vOICe is a very easy to use freeware capable of running in all sorts of computer hardware and is dedicated to people with visual impairment. By combining it with a camera for live video capturing, it can provide the user an idea of what is in front of them by producing sounds, essentially giving them the perception of space. Although not a plugin or application for audio editing and composing, the vOICe focuses on functionality by converting picture to sound creatively, while spreading awareness about visual impairment.

2.3.1. History

In 1985, Peter Meijer, after he graduated with a master's degree from the physics department of Delft University of Technology in Holland and while he was working with signal processing software and hardware of the 1990s, got the idea of making an experimental device for converting image to soundscapes in real time, which was named 'The vOICe'[11]. This complete system for capturing, representing and processing optical data consisted of a prototype computer, a camera, a pair of headphones or speakers and a monitor. In February of 1992, a scientific paper was published in the IEEE[12] Transactions on Biomedical Engineering based on The vOICe system, in which Meijer presents how the image data to sound conversion works while explaining the main reason behind the idea, which is to work as a supplementary system to human hearing or even to help visually impaired people to see by representing images as sound (Meijer, 1992). From the beginning, he visualized the possibility of a portable device or a wireless system that, with the appropriate hardware technology, would make the conversion from captured image to sound in

[11] The three capital letters in the middle mean 'Oh I See'

[12] Institute of Electrical and Electronics Engineers

real time, giving someone the perception of space without seeing in the form of augmented reality (*The vOICe - The Index Project*, n.d.).

The first version of The vOICe as a software was available in 1998 for home computers without the need for specific hardware apart from the essential headphones and camera. In 2001, an attempt was made to install the software in sunglasses with a built-in hidden camera in order to be used by the visually impaired while the exact setup for remote functionality, including a laptop, would cost $2,500 at the time (Sandhana, 2003). Later versions of smart glasses received international recognition by the scientific community with the latest model being VISION-800 (Picture 16), which is available for purchase and was used by Vadim Artsev in his entry in the Neurothlon 2018 competition, winning the first prize among several competitors, even ones using more expensive iris implants (Rehabilitation industry of Russia, 2019).

Picture 16: The VISION-800 smart glasses

2.3.2. Overview

At the time of writing, The vOICe is available to download for Windows PC, smartphones using Android and as a browser application. All versions are free. The version 2.12 of 'The vOICe for Windows (The vOICe Learning Edition)' which is compatible even with Windows 11 64bit will be used for testing purposes. The way the software works does not differ much from the other in the image reading aspect but is oriented towards a different sound output and not music synthesis.

The application gives access to the camera in order to record live image and scan it from left to right every second. The image is analyzed in greyscale and not colored unlike other programs, while at the same time there is sound output in stereo, so live

video is scanned and played according to differences in black and white like reading a page. Sound changes depending on the color intensity at each point that gets scanned. It is observed that there are more profound sounds if there is high contrast between black and white, where white areas produce steady and high-pitched white noise while black areas produce sound that fluctuates depending on the area position in the vertical axis. Audio fluctuations are also extremely precise and fast, giving users the ability to distinguish very small details such as the points of a fork. All the above can be evaluated by many published videos presenting the equipment either through the Neurothlon competition or in home use, where the user tries to avoid obstacles or pick up bright white objects. With proper training the user will get more familiar with the system in order to feel, even approximately, whatever is in front of him by hearing. The application is very user friendly and easy to work with, having quick access to some basic settings for video scanning, cursor speed, negative video, high contrast and edge enhancement.

2.3.3. Conclusions

Most of the programs analyzed in this thesis are mainly digital signal processing tools that use images as data input in order to synthesize sounds, but The vOICe differs in functionality due to the end goal being to give users the perception of space through sound. It plays its part only in the sense of converting live image to sound, therefore it cannot be compared to the rest regarding software features and updates for modern operating systems as it falls in its own category. Although different, it certainly is unique and worth mentioning.

2.4. MetaSynth by U&I Software

Metasynth is a sound design and music production studio, which works as a software synthesizer, sequencer and sample editor. Its use of a graphic editor to control and generate audio, places it among similar software which use colors, geometric shapes and sketches to interpret to audible sound. The unique resulting sounds of Metasynth have picked the interest of many artists in the music and film industry, which has built a reputation among famous electronic musicians, DJs and soundtrack composers.

2.4.1. History

MetaSynth is made by U&I Software, a software company mainly known for creating digital applications for artists and musicians. The two programs that they provide at the time are MetaSynth and ArtMatic, with the first being a sound production tool and the second being a modular synthesizer for 2D and 3D graphics and audio creation. Both applications are available only for systems running macOS. The first

version of MetaSynth was released in 1998 (*U&I Software MetaSynth - Specifications, Pictures, Prices, Links, Reviews and Ratings*, n.d.) and currently exists as version 1.4.2 from the official website.

2.4.2. Overview

The application is completely self-contained, unlike other audio workstations, as it does not have options for connecting plugin effects or instruments or even be controlled via MIDI. The interface is similar to picture editing software such as Adobe Photoshop, with the main element being the image display or drawing canvas in the middle. Using very similar digital tools to the aforementioned programs, one can paint anything imaginable for the program to convert to sound. Much like on a spectrogram, time and pitch are represented on the horizontal and vertical axes respectively. Its primary goal is sound synthesis by graphically manipulating the canvas area. The application uses additive synthesis to interpret what is drawn, or the image that it takes as an input but "the graphical functionalities […] are constrained by the additive synthesis method" and "in the end, the user draws more or less a spectrogram which limits the originality of the sounds created, a disadvantage shared with the UPIC system" (Thiebaut et al., 2008).

The interface of MetaSynth contains many different tools for the user to interact with. First of all, there is the signature composition environment called Image Synth, where paintings become sound. While MIDI can carry a fixed amount of information, that being note on and off messages or modulation envelopes with a maximum of 127 notes and 32 voices, the Image Synth can integrate thousands of voices tuned independently in different ways. Each pixel of the imported image will enable an oscillator with brightness determining amplitude and the color dictating the stereo position. For instance, yellow-colored shapes are panned in the center, red are left and green are right, while a different color like blue puts the sample on another audio layer. Also, there are five built-in synthesizers to choose from that are simple, efficient and optimized for the high voice count that the software is capable of generating (Dow, 2023).

2.4.3. Conclusions

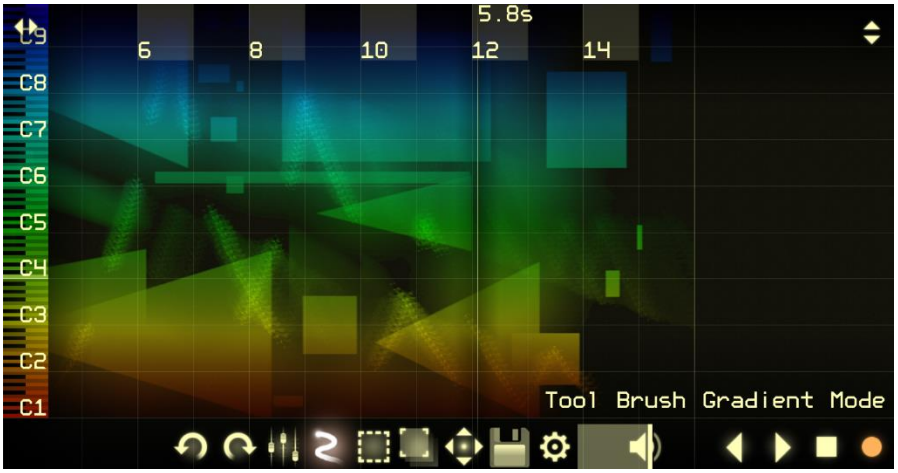
Within the immense selection of intricate digital audio workstations and audio design software, MetaSynth is able to keep its title as one of the most unique applications to exist for over two decades. Still to this day, it is used on numerous film and video game soundtracks, not only as the main software for sound design, but also as a supplementary tool for producing more extraordinary sounds.

2.5. Virtual ANS by Alexander Zolotov

The Virtual ANS is a modern version of the infamous ANS Synthesizer Murzin invented in 1958. Programmed in a custom programming language, this software emulates how the original synthesizer worked, as it also expands the options of generating sounds with added features.

2.5.1. History

Virtual ANS (an abbreviation for the Russian composer Alexander Nikolayevich Scriabin) is a portable experimental application, which simulates the operation of the ANS synthesizer, a unique electronic musical instrument made by Russian engineer Evgeny Murzin (*Warmplace Virtual ANS*, n.d.). Developed by Alexander Zolotov in 2007, it was the first attempt to resurface this historically important instrument and its means of operation. The latest version being 3.0.3 (2023) is available to download from the official website and is compatible with Windows, macOS and Linux operating systems as well as Android and iOS for mobile devices. Virtual ANS is programmed on Pixilang, a simple pixel-oriented programming language for small graphics and sound applications, created by Zolotov himself (Picture 17).



Picture 17: The user interface of the Virtual ANS

2.5.2. Overview and sound

Very much like on the ANS synthesizer, this software simulated version composes sound from images as they are scanned, just as the spectrogram operates: the vertical axis corresponds to pitch and the horizontal to time. When played, the cursor "reads" the image from left to right and each pixel generates a single tone, meaning that depending on the size and resolution of the given image, virtually unlimited pure tone

generators are available. The application has implemented features that extend the capabilities of the original analogue synthesizer which is based on. The whole interface is now an editor upon which lines and shapes can be drawn with common graphic editing tools such as different sizes of brushes, while the image is interpreted to sound at the same time (*WarmPlace.Ru. Virtual ANS Spectral Synthesizer*, n.d.). As each pixel is a sine generator, volume is handled by the pixel's brightness, while a thin drawn line will produce a sound close to the sine, a thicker one will spread on the frequency spectrum, producing a broader sound closer to noise. Besides the direct image to sound conversion, there is also the option to do the opposite and convert sound into an image. By connecting the application to a microphone or loading a WAV file, every sound can be drawn onto the interface. In the end, users are provided with plenty of settings to tweak and play with, offering them unlimited different sound results. There are also some basic settings for adjusting ADSR[13], multiple drawing layers for editing, mixing and saving the sound result. The sound output is close to 1960s ambient electronic music and very atmospheric, reminiscing the sound used by the original ANS synthesizer at the time to accompany movies like Solaris (1972) and Stalker (1979) by Russian composer Eduard Artemyev.

2.5.3. Conclusions

The Virtual ANS application pays a tribute to the analogue system that it is based on, by not only simulating its sound output in a very similar way but also by implementing modern features that could not possibly exist at that time. It is a great example of a simple and effective audio creation tool that utilizes images, is free to use and has resurfaced the inner workings of the historically important ANS synthesizer. Many users compare it to the image to sound conversion and editor that exists in MetaSynth, although being a cut down version, it accounts for a well-made free alternative.

[13] Attack, decay, sustain, release

3. Pix2MIDI

In this chapter, the creation of Pix2MIDI will be examined, which is an original standalone application made in the Max environment capable of utilizing data from images, in order to control VST instruments and convert them to sound. Every feature set and fundamental protocols that this application is based on, as well as its connection to musical aspects, will be explained thoroughly. Moreover, future improvements will be discussed as well as the possibility of a dedicated hardware implementation.

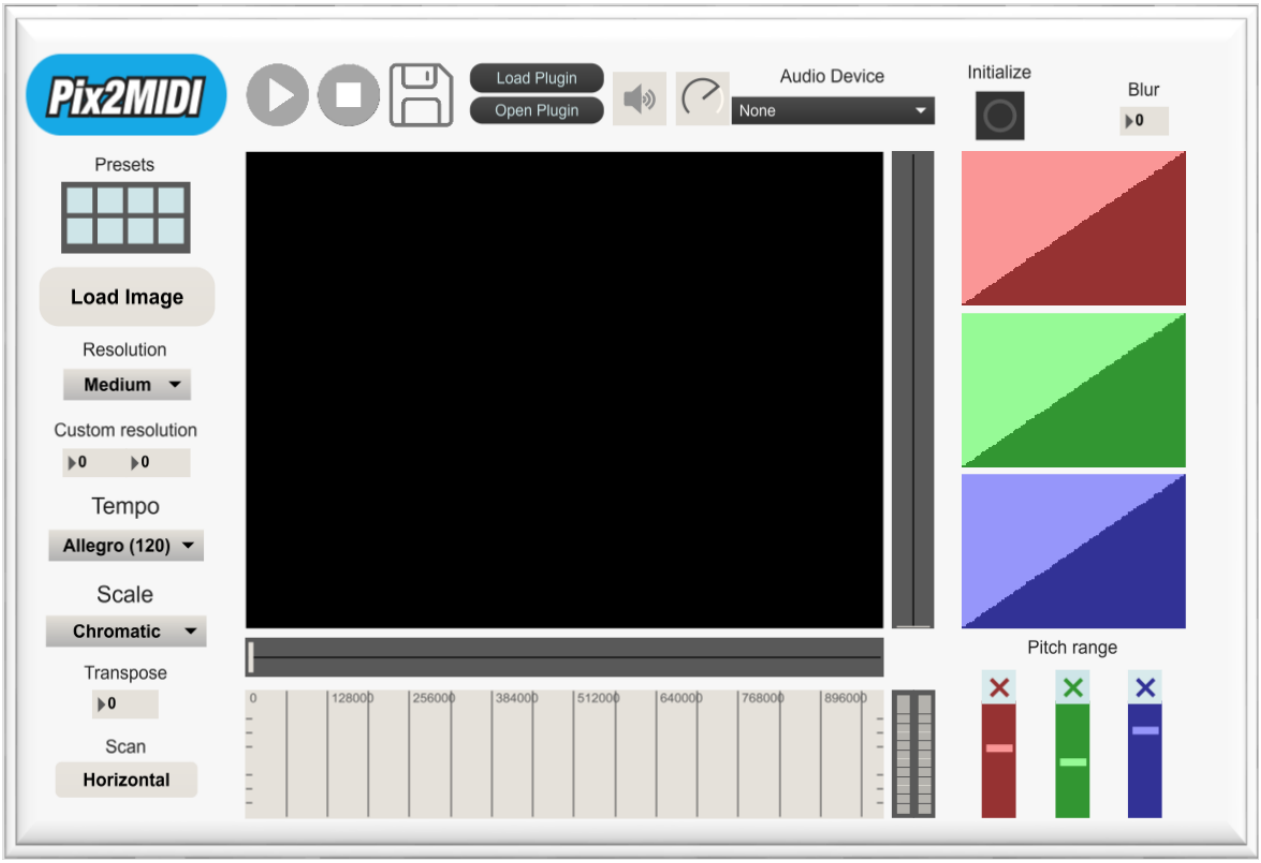
3.1. Overview

The initial inspiration behind the development of the application does not differ from what the other creators were inspired by; Alexander Scriabin's groundbreaking connection of music and colors played a major role in the conception of the idea. Moreover, the complexity of graphic scores in contemporary music and the simple notion that shapes can represent sounds, sparked the need to explore the possibility of taking an image or a photograph and merely seeing – or more appropriately hearing – what kind of "music" it can make.

Pix2MIDI is a data managing application that exports MIDI events using information derived from pixel characteristics of an image and it is built using the Max visual programming language by Cycling '74, as a standalone software for Windows PC. This visual coding environment, called a 'patch', is programmed to break down the three main colors of pixels from an imported image and read their values in different orientations, which are then mapped accordingly in order to control and manipulate music characteristics via MIDI messages, such as note pitch, velocity and tempo. Through a variety of settings for editing musical aspects and the image itself as differently colored images output different values, it provides endless possible sound outcomes boosting user's creativity. Pix2MIDI can ultimately be thought of as a MIDI controller, as it does not produce sound by itself, but has to be connected to any virtual instrument that receives MIDI events as an input to be able to output an audio signal. For convenience purposes and in order for the application to be ready to use and play sounds without needing any initial technical knowledge from the user, it includes 'Dexed' by Digital Suburban, a free VST plugin synthesizer preloaded in the standalone version.

3.1.1. User Interface

Pix2MIDI is built directly from the Max environment as a standalone application capable of running on Windows compatible PCs. The user interface is made to be easy to read by including drop-down menus, colored sliders and buttons. In the top section there are buttons for basic controls such as: play, stop, save, plugin management and audio device control. On the left-hand side, there is every setting that controls the music and scanning options. These include menus for resolution, tempo, scale, transposing and scan orientation, as well as the preset menu and 'load image' button. On the right-hand side, the settings include the color editing sliders, initialize button and blur value, as well as the colored pitch range sliders. Finally, the display which hosts the image to be scanned is in the middle (Picture 18).



Picture 18: The Pix2MIDI user interface

3.1.2. User manual

Right after launching the Pix2MIDI standalone application, it is necessary to ensure that the proper sound device is selected. The 'Audio Device' drop-down menu gives

access to every sound device installed on the computer in order to select the preferred one. Then, to turn on audio, click the speaker icon and it should light blue when it is turned on. Next to the speaker icon there is a dial that controls the master volume. Note that after configuring these settings for the first time, when launching again, the application stores the selected sound device and enables the audio output automatically. As mentioned before, Pix2MIDI works as a MIDI controller and it must be connected to a plugin virtual synthesizer to produce sound. By clicking the 'Load Plugin' button, a file explorer window appears for the user to select the plugin device to be used. For convenience purposes, the application comes pre-loaded with a free VST synthesizer called 'Dexed'. By clicking the 'Open Plugin' button the user can access the control panel of the loaded plugin in order to adjust the individual settings that control the sound quality (Picture 19).

Picture 19: The control panel of Dexed

Once configuring these initial settings, the user can effectively load an image to be scanned and analyzed in terms of color intensity. Any other settings on the interface

are focused on altering the sound output through editing the image colors or adjusting the music settings. Click the 'Load Image' button for another file explorer window to appear in order to select and load the image to be converted to sound. Then, the loaded image appears on the center screen and it is ready to be edited in terms of color intensity. The three colored panels on the right side, which are composed of many small sliders can be adjusted by clicking and dragging the left mouse button. As the sliders are moved up or down the color of the image changes, which also determines the final sound pitch. Also, above the colored panels there is the setting for blurring the image, as well as the initialize button, which restores the image to its initial state.

The loaded image can also change in resolution via a drop-down menu in real time without loading it again. As it will be explained in the following chapters, the resolution and therefore the image size, determines the duration of the final sound output. The custom resolution dimensions only need to be provided with a number when 'custom' is selected in the resolution menu. Additionally, the musical aspects of the sound can be configured by selecting the preferred tempo and musical scale from the menus, or by entering a number in the 'Transpose' number box, which changes the tone according to the value provided in semitones. In the 'Pitch range' part of the interface there are three colored buttons and sliders. These correspond to three separate tracks, as the values of each color of the image produce a sound on their own. You can choose which individual color track is enabled or disabled by clicking on the colored X button and adjusting the sliders below to configure the pitch tone range. One of the most important parts is the scanning orientation of the image. Click the interactive button to switch between horizontal and vertical scanning.

Finally, after every setting is configured to each user's preference, click the play button to start the image scanning and audio playback. If clicked again, the play button works as a restart and in order to stop, just click the stop button. When the audio playback is finished you can save the audio sample on the computer by clicking the floppy disk icon.

3.2. MIDI standard

MIDI, which is an acronym for Musical Instrument Digital Interface, is a technical standard that is used to store and transmit music data between digital devices. It is a communication protocol for computers, digital musical instruments and plenty of other hardware devices used among them to "speak" to each other. It was developed in the early 1980s by the synthesizer company Sequential Circuits after realizing the need for a global standard because at that time electronic musical instruments of different manufacturers had their own proprietary standard to get synchronized. It

was first implemented in 1983 by the release of two historical instruments, the Roland Jupiter-6 and the Sequential Circuits Prophet 600. MIDI does not transmit audio signals produced by instruments in order to communicate between them. Instead, it sends information via MIDI messages which then get decoded by the receiving device into sound, much like the punched paper sheet music of a player piano. These messages can include data about pitch, velocity and aftertouch pressure, volume control or channel, basically every main feature on a synthesizer keyboard. It also can send system control information such as messages for program change, instrument sounds and time code for synching the time clock. The MIDI protocol takes values ranging from 0 to 127, in that manner every instrument sound and every system control setting apply to a specific number, as well as the amount of said setting, for example volume has the same range. The MIDI Control Change list contains every number that corresponds to a setting while the MIDI Program Change list does the same for instrument sounds. All this data can be compressed, formatted and sent to electronic instruments as a single message which is called a MIDI event.

Pix2MIDI is very much based on this industry standard as it outputs MIDI events and works like a MIDI keyboard controller. As the Max environment is optimized for MIDI data management, this application takes advantage of formatted MIDI events to control pitch, velocity, channel and volume on the connected VST instrument.

3.3. Scala format

Music scales in the application are handled by a new feature added in the Max environment on version 8.5 (2022). The Scala Format, created by Manuel Op de Coul, is a software tool capable of managing music scales saved in a single file, in order to experiment with different musical tunings and load them into various hardware and software. It offers complete creative freedom, as scales can be assembled from scratch with whatever number of notes and the distance from one to another can be defined with the precision of one cent. Having in mind that a single semitone is one hundred cents, this format transcends beyond the 12-tone equal temperament, being capable of making any scale imaginable such as just intonation ones, historical temperaments and even microtonal and macrotonal. (Op de Coul, n.d.). Max supports this format via the objects ftom, ftom~, mtof and mtof~ which offer basic functions to convert MIDI notes to frequency and vice versa. There is also an archive of more than five thousand different scales built in these objects for the user to recall, but for this application, the Scale Workshop web platform will be used to construct custom ones.

The Scale Workshop provides the ability to view, edit and export music scale data in plain text or even in extremely specialized formats compatible with specific

equipment (*Scale Workshop · Xenharmonic-Devs/Scale-Workshop Wiki · GitHub*, n.d.). A music scale can be analyzed in two ways: by measuring the distance between each note in cents or by defining the ratios of musical intervals according to the Pythagorean monochord. The equal division of a musical scale in intervals with fixed distance between them became established in western music by Johann Sebastian Bach from a collection of works he published in the 18th century named 'The Well-Tempered Clavier', containing pieces written in every standard major and minor scale. Thus, in this system known as equal temperament, a musical octave consists of twelve semitones placed 100 cents apart. In the Pythagorean monochord, musical intervals originate from pressing a string in certain points according to specific ratios known as just intonation. For example, if a string is pressed in the two thirds of its length (3:2 ratio) and then plucked, a note distanced by an interval of a perfect fifth (seven semitones) above the tonic will be heard. This can be validated using any fretless stringed instrument like a violin. In Scale Workshop, notes of a scale are inserted in either of the two ways, by cents or ratios, while additionally, the center base frequency of 440Hz which corresponds by default to MIDI note number 69, can be shifted. With this web application anyone can create any musical scale imaginable and export it in a variety of formats, although for it to be compatible with the Max environment it needs to be exported with the extension SCL for Scala scale.

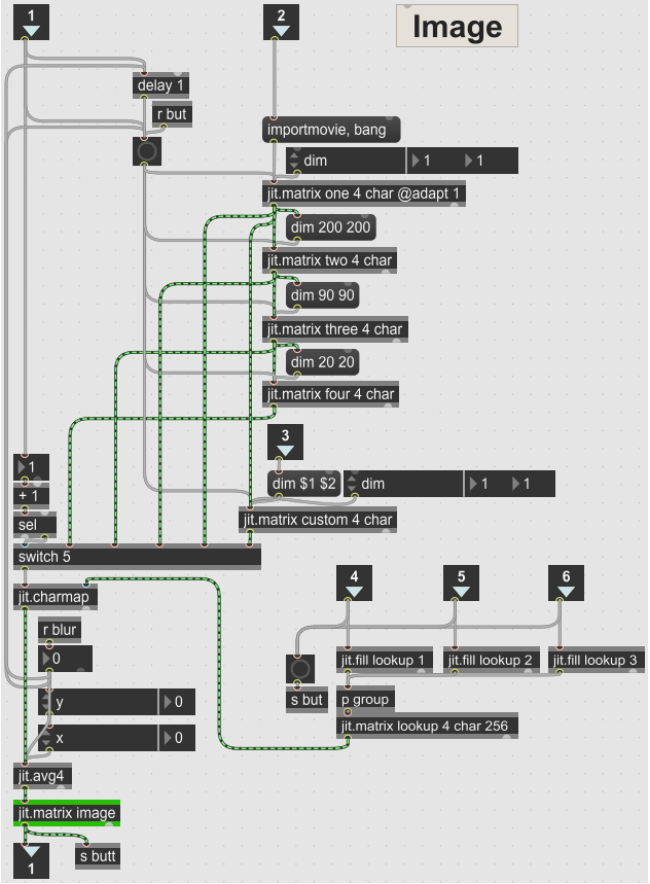
3.4. Patch analysis

In this part, there will be a thorough examination of each section of the patch in order to explain how the image is loaded, edited and ultimately converted to sound, how the data conversion is done and finally what features are available for the user to interact within the user interface, while also focusing on musical concepts. There are two 'sub-patches' which contain the code for image importing and the scanning loop, while the rest of the editor contains the audio channels, data routing, recording as well as some of the music settings.

3.4.1. Image importing and editing

The first step in using the application is to import an image by pressing the 'load image' button after setting the preferred resolution. When importing an image, the system loads it in five different two-dimensional matrices at the same time so that every given resolution is available to be selected at any time, without loading again. The system automatically adapts the image to square shape to ensure consistency when changing the scanning orientation, which will be explained later. There are options for low (20x20), medium (90x90), high (200x200), as well as original image size without adapting and custom resolution. When custom resolution is selected, the

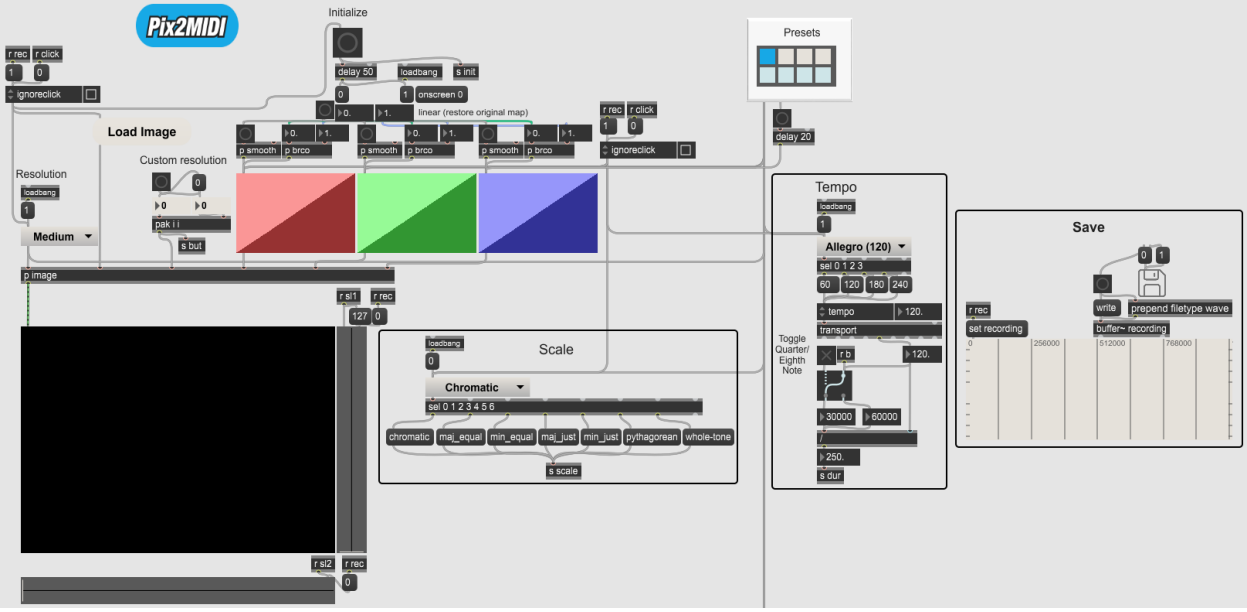
exact dimensions for x and y axis need to be set, with the image being constructed in real time. Apart from resolution, there are some basic options for editing the loaded image. The blur factor, when provided by a number, shifts a copy of the image vertically and horizontally making it blurred. These parts of the code are enclosed in the ‘p image’ sub-patch (Picture 20).



Picture 20: The 'p image' sub-patch

There are three colored interactive panels, which are used to alter the color intensity by increasing or decreasing the value of each color in every pixel area. For instance, when the user interacts with the red color panel and drags the slider points higher - every color panel is composed of 256 individual sliders, each one mapped to a specific pixel area - the red color in each pixel is increased, resulting in a more red image. By default, the image is loaded and analyzed in the most common pallet for computer graphics which is the RGB color space. The panels colored red, green and blue, control the intensity of each color in the image so the user can experiment with different color values and combinations. The above-mentioned editing options play a significant role

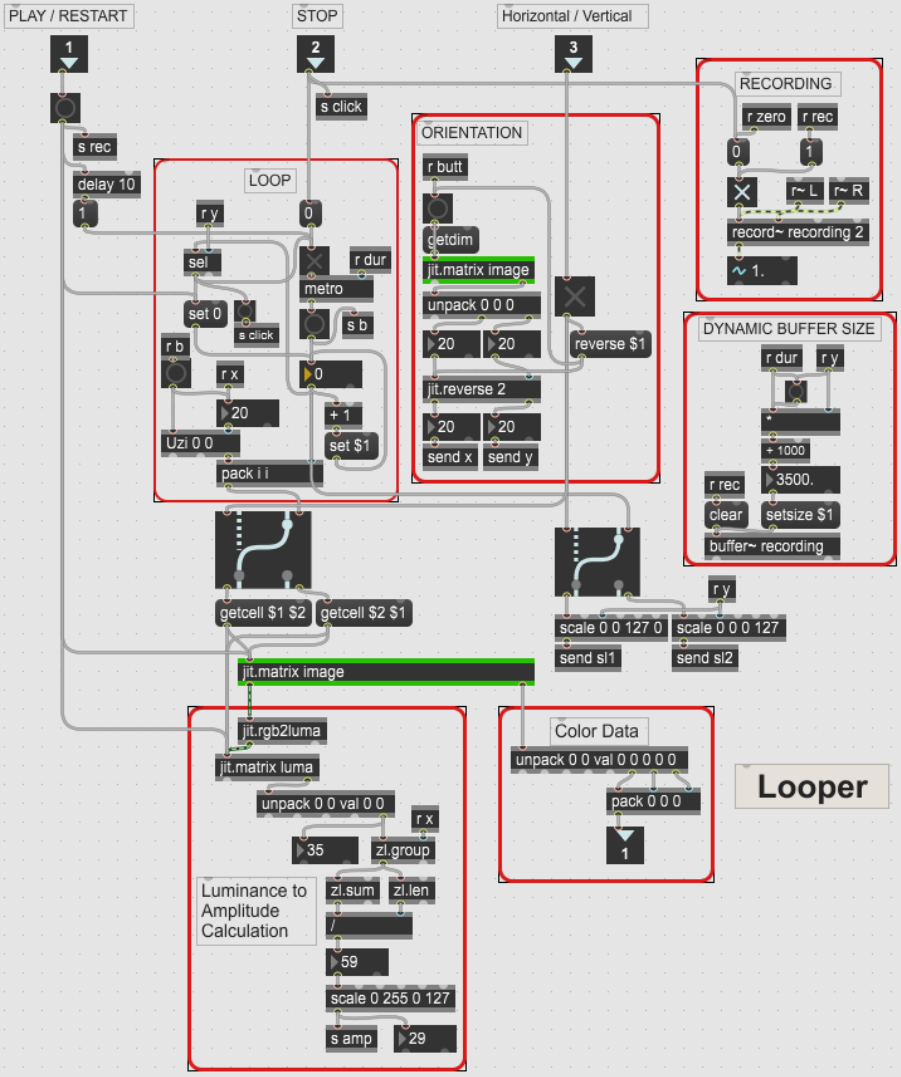
in sound output. Finally, by pressing the initialize button every setting in the editor is reset to the initial state (Picture 21).



Picture 21: The first half of the Pix2MIDI main patch

3.4.2. Scanning loop

Settings referring to image scanning and audio recording exist in the 'p looper' sub-patch (Picture 22). In order for the application to convert the image to MIDI messages while scanning it, it needs a way to read and output the data of each individual pixel. The loop is constructed by selecting each image cell one at a time - row by row and line by line - until the whole image is scanned, with a rate defined by the tempo. That way, as the tempo increases, the image is scanned faster and so the length of the composition shortens. Although, the end duration is also affected by the image dimensions, as a higher resolution one contains more pixels to read. This is the reason why the loaded image is scaled to fit a square, because even if the orientation is changed, the audio output duration remains the same. Be aware that when loading an image with a different aspect ratio than 1:1, the results will definitely vary. By default, the application is set to scan the image vertically, but with the press of a button the orientation can change to horizontal, switching the x and y axes.



Picture 22: The 'p looper' sub-patch

3.4.3. Music settings

The first module in the music settings controls the tempo or, in other words, the speed of image scanning. There are options referring to common tempo marks found in standard music notation and counted by beats per minute, specifically: Adagio (60bpm), Allegro (120bpm), Presto (180bpm) and Prestissimo (240bpm) (Picture 21).

As mentioned previously, scales are automatically loaded in scala format through a custom object and are not imported from the Max scale archive. The *pdm.scala.parser* object is a custom made patch by Philip Meyer, available for free on his GitHub page, created specifically to make scala format files work with the compatible objects

without the need to recall them from the archive (Meyer, 2023). Through that, every custom scale, which was exported in scala format from Scale Workshop, can now be loaded efficiently. Below, the custom-made scales will be analyzed by listing their ratios or the distances of individual notes in cents which were used to construct them. When talking about distances in cents, each value corresponds to the distance from the tonic. Every scale, besides the two chromatic ones, has the note A at 440Hz as their tonic:

- Chromatic
  A standard twelve-tone chromatic scale in cents:
  100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200.
- Major equal
  A major scale in equal temperament in cents:
  100, 300, 500, 600, 800, 1000, 1200.
- Minor equal
  A minor scale in equal temperament in cents:
  100, 300, 400, 600, 800, 900, 1100, 1300.
- Major just
  The same major scale in just intonation in ratios:
  1:1, 9:8, 5:4, 4:3, 3:2, 5:3, 15:8, 2:1.
- Minor just
  Also, the same minor scale in just intonation in ratios:
  1:1, 9:8, 6:5, 4:3, 3:2, 8:5, 9:5, 2:1.
- Pythagorean
  A chromatic scale based on ratios of the Pythagorean monochord:
  256:243, 9:8, 32:27, 81:64, 4:3, 1024:729, 3:2, 128:81, 27:16, 16:9, 4096:2187, 2:1.
- Whole-tone
  A scale containing only tones comprised of six notes in cents:
  100, 300, 500, 700, 900, 1100.

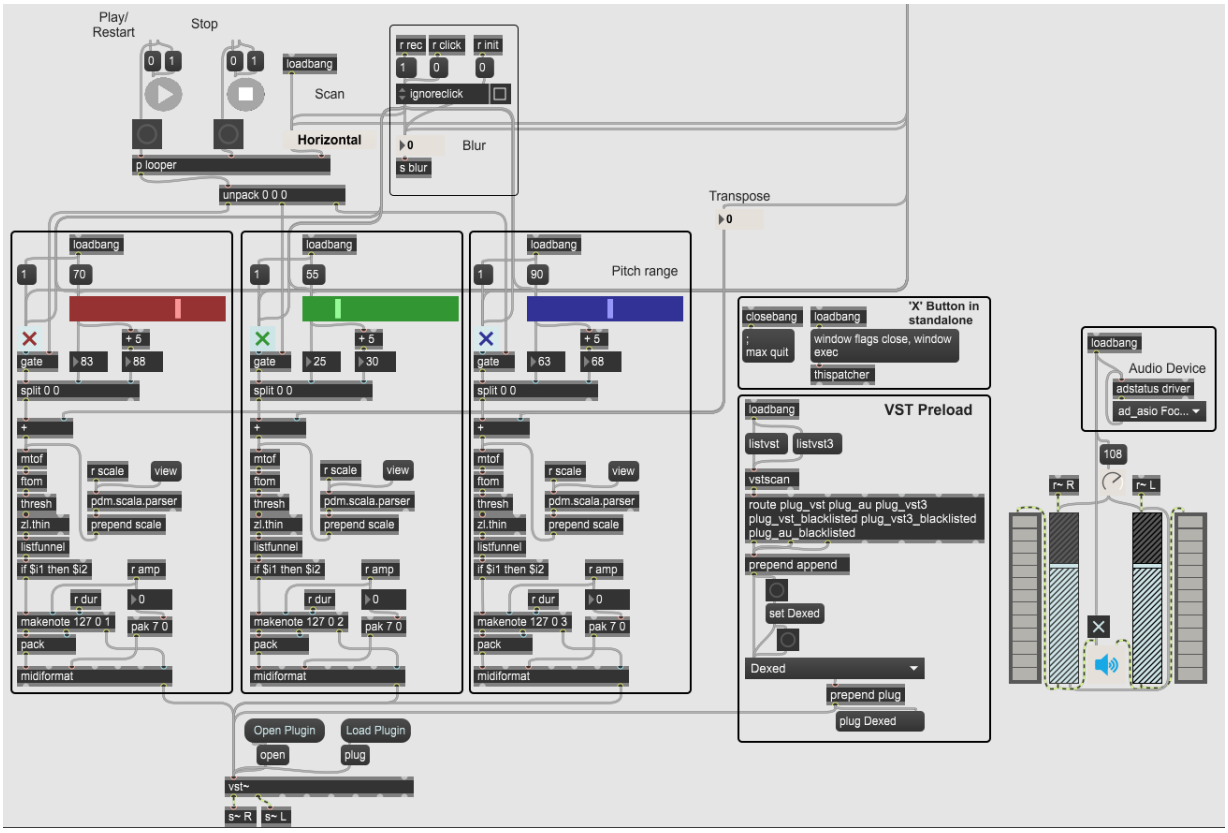
Additionally, in order to include more possible tonalities there is an option for transposing the audio result by steps of a semitone from minus one to plus one octave.

3.4.4. Data extraction and routing

Pix2MIDI does a very good job at managing and routing data from individual colors to MIDI messages, in other words, it does data mapping. As mentioned, the application scans the whole column or row of the image - depending on the selected orientation – one at a time - and outputs the values of every pixel for each color. The extracted numbers represent the intensity of each of the three colors in the RGB color space inside those pixels, which range from 0 to 255. These values are then scaled to

fit the MIDI note range which is between 0 and 127, and are loaded into three separate tracks (Picture 23). Note that, as the full length of the column or row is analyzed in every instance of time, the audio tracks receive as many values as the set resolution of the image simultaneously. Then, in order for the user to be able to select the preferred pitch height for each track, the program isolates a range of five numbers from the color data. If any extracted color value falls between the set pitch range numbers, it then gets reconfigured to fit in the selected musical scale according to the constructed scala format file.

On the other hand, the note amplitude value originates from how much illuminated each column or row is. After loading, the image is also converted to black and white in order to calculate the luminance. While it is scanned, a single number that represents the luminance value for the column or row is extracted besides the color data (Picture 22). This specific value determines the amplitude of the output sound. Finally, after any duplicate numbers are removed, the data is formatted as MIDI events, which include the pitch and amplitude characteristics according to the scaled color values.



Picture 23: The second half of the Pix2MIDI main patch

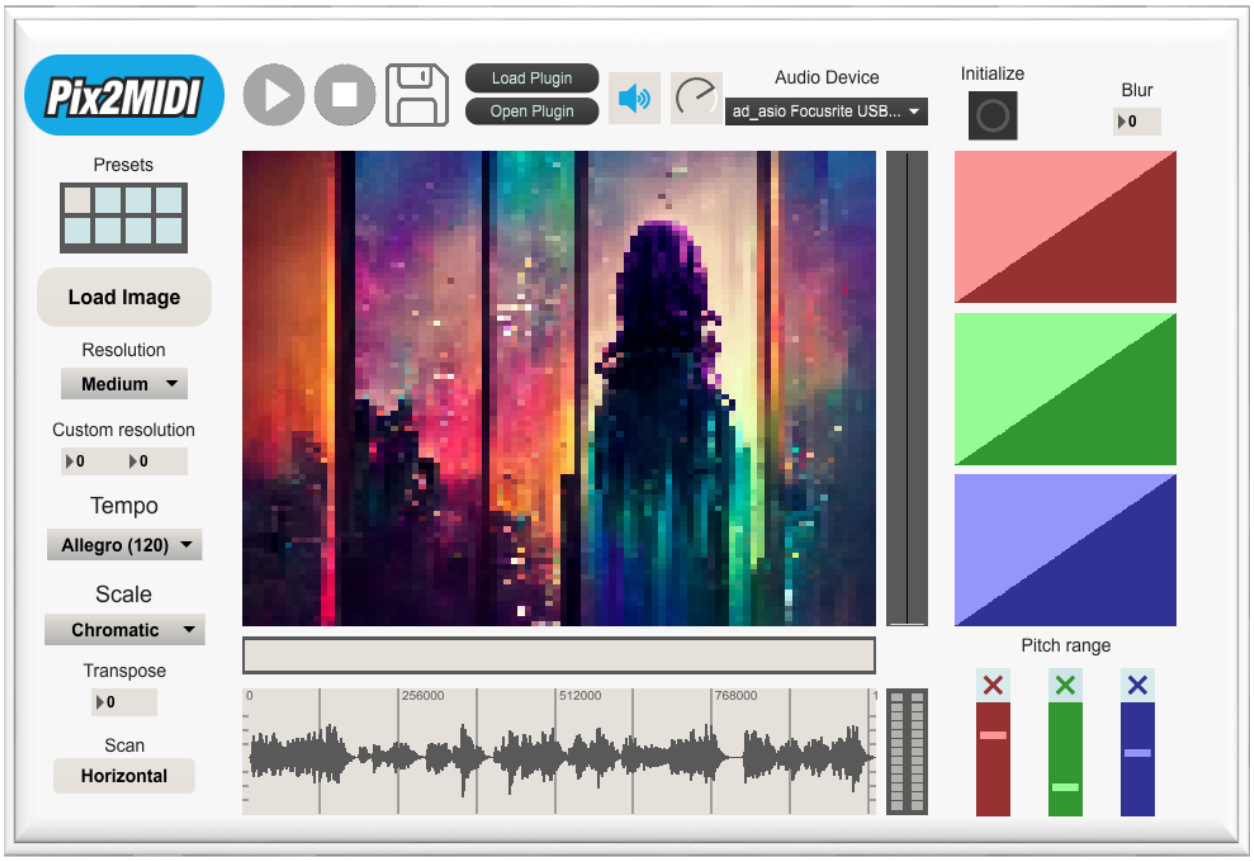
3.5. Audio examples

In this chapter, four examples are presented, that include an image along with the waveform of the audio output that Pix2MIDI provided. These picture examples differ in color variety and contrast in order to demonstrate the capabilities of the image to audio conversion. The audio results in each case will be discussed and analyzed for their efficiency in representing musical characteristics. The loaded presets in the application provide quick access to the settings used to produce the audio samples, so that anyone can replicate these examples after loading the corresponding images. Finally, some of the examples use a different audio preset loaded in the 'Dexed' virtual synthesizer. The first example uses the 'LAURIE' preset, while the rest use the preset named 'PHAROH' from within the interface of the virtual instrument.

3.5.1. Example 1 – Girl

The first preset includes the settings that result in the audio sample provided for the 'Girl' picture. This AI generated image example contains a large variety of colors which are spread across the color spectrum including bright and dark areas mostly on the lower part. It is loaded in medium resolution and scanned horizontally to demonstrate accuracy and how the application responds to thin colored columns that are bright or dark.

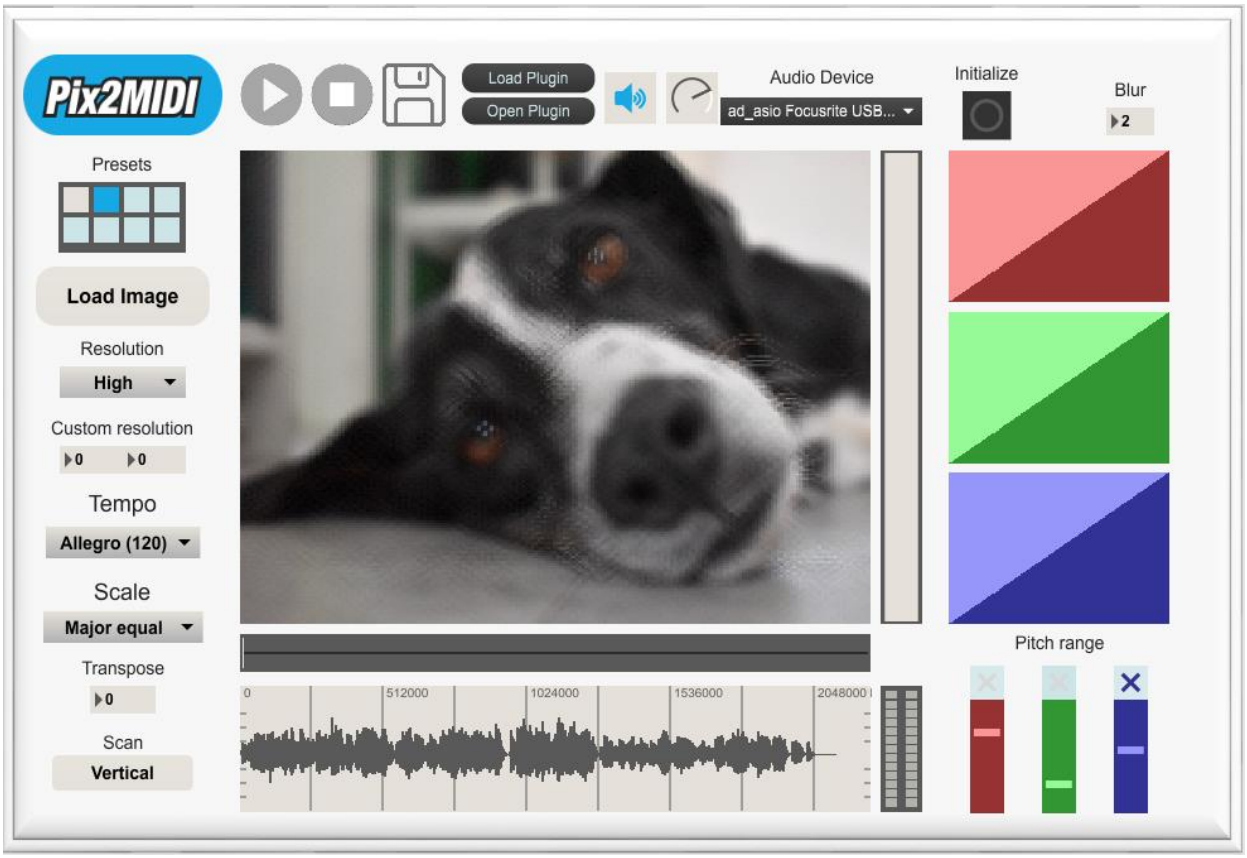
On the 3.0 second mark - where there is the first orange-red column across the whole vertical length of the image - there is a peak in audio volume due to the high brightness. The difference in brightness is significant before and after this pixel column and the sound fluctuation becomes noticeable immediately. The same happens on the 19.0 second mark where there is a larger area of bright pixels. Also, there are two major points that resemble musical rests on the 4.0 and 18.0 second mark. The first one aligns with the dark column area right after the volume peak, while the second one, although being on a bright area, does not output audio because the color hues are not in the span of the selected pitch range (Picture 24).



Picture 24: Example 1 – Girl (Guliyev, 2022)

3.5.2. Example 2 – Dog

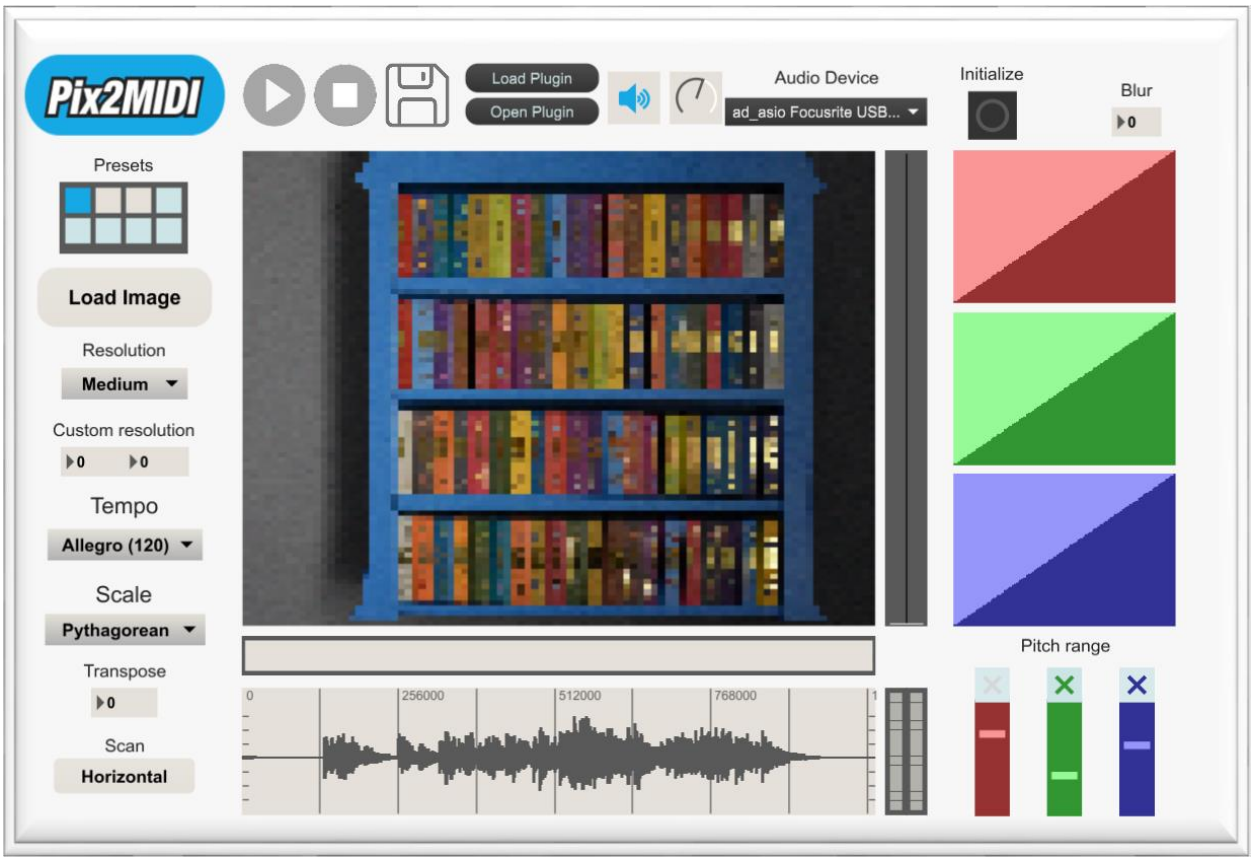
In this example, the picture does not have many colors. It is mainly composed of black and white, but in the application only the audio track that outputs audio deriving from blue color data is enabled. Although the blue color is not evident in the image, the pixels contain a small but sufficient amount to be used as input data. The image is scanned vertically this time to demonstrate the amplitude fluctuation in the darker areas. Specifically, it can be heard that around the 29.0 second mark, the audio result is quieter up until the scanning is finished (Picture 25).



Picture 25: Example 2 – Dog

3.5.3. Example 3 – Bookcase

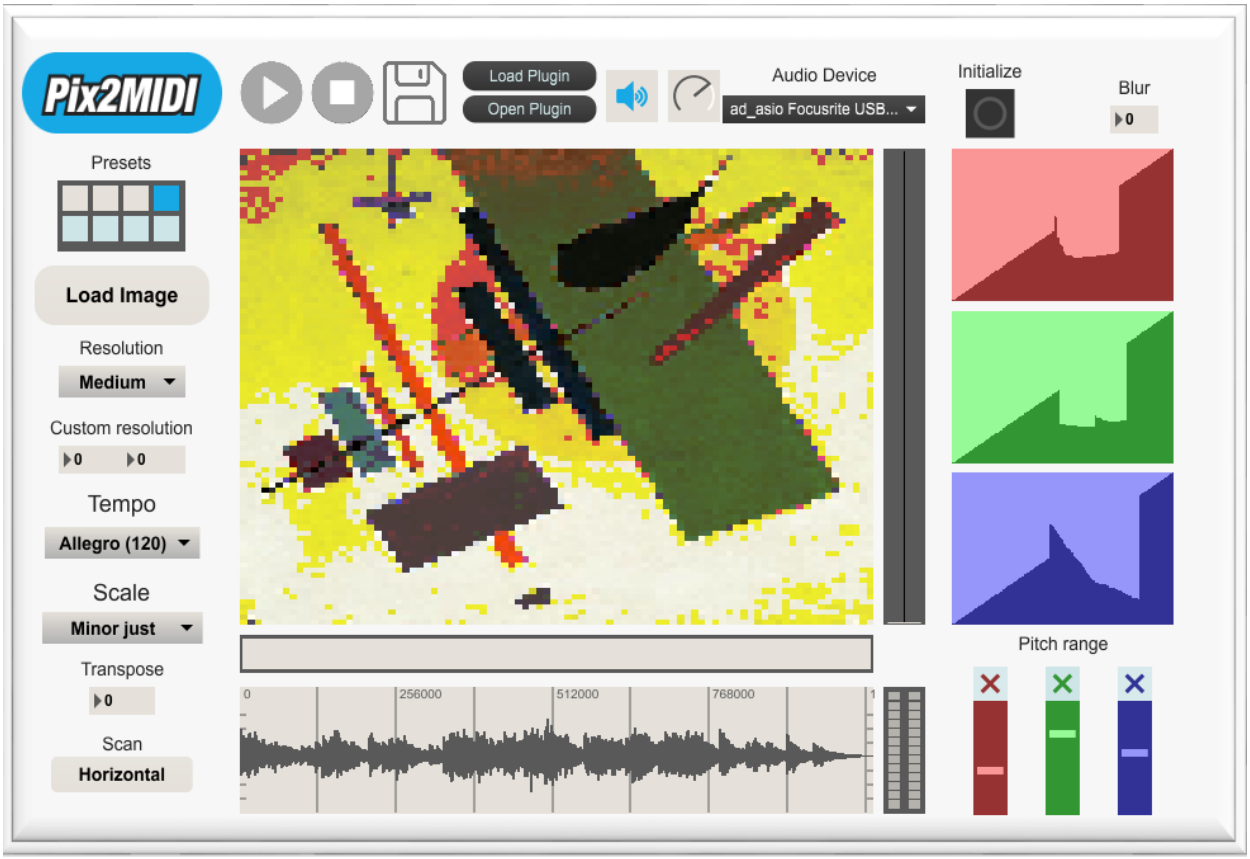
This third example is used to demonstrate that images with dense and colorful areas do not always come with a well-constructed and musically meaningful sound result. In this preset, two of the three audio tracks are enabled to manage image data, the green and the blue. Besides that, there is almost no audio output in the first and last three seconds because of the absence of color outside of the area of the bookcase. Regarding the middle part of the image where the bookcase is, although being colorful, it sounds like musical chaos with many clusters of notes being played continuously. This example proves that any given image needs to be configured, either by adjusting the sliders that control the presence of the individual colors or by changing some of the music settings in order to have a meaningful audio result (Picture 26).



Picture 26: Example 3 – Bookcase

3.5.4. Example 4 – Painting

In the last example, the application receives the painting 'Number 55' by the avant-garde artist Kazimir Malevich (1879-1935) as an input. For the purpose of that demonstration, the colors of the loaded image are edited by configuring the three colored panels on the right side. From the start of the audio sample, the presence of mostly the yellow color gets picked up by the third audio track, which manages the presence of blue color, and through the whole length of the composition it provides a continuous repetition of notes. At the same time, because there is not much red or green, the other two tracks produce less notes which are present in the middle of the audio sample. These notes, being higher and lower in their pitch range, provide the impression of question and answer between voices. Also, the minor selected scale becomes more prominent through musical intervals that get created (Picture 27).



Picture 27: Example 4 – Painting

3.6. Future vision and improvements

As mentioned previously, Pix2MIDI is essentially a MIDI controller as it is programmed to generate MIDI events and needs to be connected to a digital instrument in order to produce sound. Right now, there is only the option to connect digital instruments in the standalone application. For the capability to connect hardware synthesizers in order to be controlled by picture data, the application must run through the Ableton digital audio workstation. Ableton has native support for anything that is built using Max and it can run scripts of code directly. As a result, through there, a physical instrument can be connected and controlled by Pix2MIDI. But a major upgrade would be to implement the software onto a self-contained, dedicated hardware device, on which the user can then connect any physical synthesizer.

With the Max version 8.5, the company integrated the RNBO[14] add-on in the programming environment. RNBO enables the ability to efficiently export any creative work done in Max and deploy it to various software and hardware devices, while also being capable of exporting C++ and JavaScript code for custom desktop and web applications. There is a list of new compatible objects in the environment which operate just as regular Max objects that already exist, but with the added feature to be ready and pre-configured to export to dedicated devices. One of the most well-known and open-source hardware devices is the Raspberry Pi; a small but powerful mini-computer which can be configured with custom software and a variety of hardware components such as physical knobs, sliders, LEDs and screens. Through RNBO there is the option to export the Max code and run it on one of these mini-computers.

Therefore, the components of Pix2MIDI that the user needs to interact with can be installed on a Raspberry Pi computer as physical parts. This would include a screen, buttons for entering the scale or resolution settings, sliders for pitch range and knobs for volume control, as well as the USB, Ethernet and audio ports that already exist on the mini-computer. But the end goal is for the Pix2MIDI as a MIDI controller to connect to hardware synthesizers, so it also needs a 6.25mm audio jack for that. Then the user would import the image file via USB and select it through a file menu, configure any preferred settings and plug the device in a synthesizer. This setup does not seem like an easy plug-and-play task to be deployed directly from RNBO to Raspberry, as it needs custom made software to be programmed to work as a file explorer and menu configurator.

The aforementioned configuration could provide the users with access to so many creative ways of controlling their own hardware MIDI instruments. From sending MIDI events to multiple devices at once, to producing looped sequences with data from images which are displayed on a live show at the same time, the artistic possibilities of the setup could be endless.

[14] RNBO standing for Rainbow

Conclusion

This thesis aimed to explore and examine how the concept of using pictures as a means of controlling and manipulating sound has evolved in the 20th century through hardware implementations, to the present day in the form of software applications. It became evident that the artistic approach of the famous Russian composer and inventor Alexander Scriabin, who lived during a period of many changes in art and technology, was the starting point of several ideas around sound and colors. As Scriabin's synesthetic perception described the idea of sound represented in different color hues, inventors of the early 1900s created numerous devices that took the form of preconfigured or interactive electronic instruments, which incorporated colored materials to be interpreted as soundwaves. We explored some of the most influential and well-documented inventions from artists and engineers, that apart from being used from contemporary composers to write music for the film industry, they set the basic principles and techniques of converting pictures and colors to sound, for the more complicated software versions that were developed in the following decades. Additionally, we also found a similar connection of how music is represented through the use of colors, shapes and visual symbols in the contemporary style of graphic notation, by examining three defining examples of avant-garde compositions by John Cage, György Ligeti and Krzysztof Penderecki. Drawing from these examples, in the second chapter we explored how the concept of using data from images to convert them to sound and music is implemented in the modern age of technology through software applications. We examined commercially available applications released in the last 20 years in terms of the techniques of incorporating images and the methods they used to essentially convert them to sound, as it became obvious that every single one of them was influenced by the aforementioned technological innovations of hardware electronic instruments of the last century. The main goal of the thesis was to use the knowledge I gained from analyzing the hardware and software implementations, in order to create an original standalone application programmed in the Max visual coding environment.

In order to realize the original idea, a lot of thorough research was done and I was impressed to see all the earlier efforts of great artists/creators. It actually transformed my views on my project and made way to more complex ideas. Admittedly, though at first I believed I had a clear image of what I wanted to make, I found myself constantly changing and adding more and more to the patch as I explored the objects in the Jitter field of Max, that are in charge of the image and video features of the program. Additionally, due to the nature of the application - dealing with the color data produced in order to create a MIDI event - it was imperative to learn how to efficiently control the values via data routing. Having

failed to do so, it would have resulted in many lost hours of finding my way through individually mapping values to certain musical aspects.

Taking into consideration that MIDI is a protocol used to transfer particular musical characteristics between electronic devices, it was necessary to incorporate musicality to the sound result through programming in Max. For this reason, Scala Format was used, which is a newly added feature in the toolset of Max. In order to have the desired result, the need to create custom musical scales arose - which became possible by using a dedicated software - and then the use of a custom-made Max object made it possible to implement them in the patch. The difficulties were not significant enough to create major obstacles in the way, but actually they created challenges that helped set milestones throughout the programming period.

If a different approach were to be taken, I would have started by envisioning the desired sound outcome first. That is to say, I would have focused less in the programming phase at first and dedicated a lot more thought into the possible music that could be created through the use of the application. It goes without saying that the technical mind overtakes the artistic approach and thus restricts the flow of ideas. Though it would be backwards thinking – to think of the result first and then how to get to it – I believe it would have notably affected the progress of the application.

Pix2MIDI is a tool to an artist's hands that could create something unique; a visual artist could import their images and select their preferred sounds in order to create an audiovisual performance. It provides the opportunity to connect vision to hearing and possibly create something that no one else can replicate thus making it absolutely special. Moreover, were it to be taken a step ahead and with the proper tuning, it could mean a whole new area of the film and music industry; scenes and colors could be accompanied by their own sounds, making the music score of each video or film all the more astounding.

Bibliography

Campen, C. V. (2010). *The Hidden Sense: Synesthesia in Art and Science*. MIT Press.

Collopy, F. (2000). Color, Form, and Motion: Dimensions of a Musical Art of Light. *Leonardo*, *33*(5), 355–360.

Darter, T., & Armbruster, G. (1984). *The Art of Electronic Music*. W. Morrow.

Hidalgo, A., Ipinza, C., & Constanza. (2016, September 23). *Cartografiar lo intangible, El Sonido.*

Jewanski, J., Day, S. A., & Ward, J. (2009). A Colorful Albino: The First Documented Case of Synaesthesia, by Georg Tobias Ludwig Sachs in 1812. *Journal of the History of the Neurosciences*. https://doi.org/10.1080/09647040802431946

Kreichi, S. (1995). The ANS Synthesizer: Composing on a Photoelectronic Instrument. *Leonardo*, *28*(1), 59–62. https://doi.org/10.2307/1576159

Levy, B. R. (2006, May 2). *The Electronic Works of György Ligeti and their Influence on his Later Style*. https://www.semanticscholar.org/paper/The-Electronic-Works-of-Gy%C3%B6rgy-Ligeti-and-their-on-Levy/662a9830ba04dd810cd6784535ad7494e759309a

Mattis, O. (2002). CCMIX Paris: Xenakis, UPIC, Continuum [Review Article]. *Computer Music Journal*, *26*(4), 96–99. https://doi.org/10.1162/comj.2002.26.4.96

Meijer, P. (1992). An experimental system for auditory image representations. *Biomedical Engineering, IEEE Transactions On*, *39*, 112–121. https://doi.org/10.1109/10.121642

Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology*, *103*(1), 1–15. https://doi.org/10.1348/000712610X528305

Thiebaut, J.-B., Healey, P., & Bryan-Kinns, N. (2008). Drawing electroacoustic music. *International Computer Music Conference, ICMC 2008*.

Xenakis, I., Barrett, R., Bourotte, R., Couprie, P., Delhaye, C., Despres, J., & Estrada, J. (2020). *From Xenakis's UPIC to Graphic Notation Today* (P. Weibel, L. Brummer, & S. Kanach, Eds.). Hatje Cantz.

Internet Bibliography

2CAudio—Kaleidoscope | It's A Trip. (n.d.). Retrieved June 17, 2024, from https://www.2caudio.com/products/kaleidoscope#_overview

Aleksandr Scriabin | Russian Composer, Impressionist, Mysticism | Britannica. (2024, April 23). https://www.britannica.com/biography/Aleksandr-Scriabin

CCMIX Paris: New Electroacoustic Works – Mode Records. (n.d.). Retrieved June 17, 2024, from https://moderecords.com/catalog/098_9ccmix/

Crab, S. (2014, February 19). 'UPIC system' (Unité Polyagogique Informatique du CEMAMu) Patrick Saint-Jean & Iannis Xenakis, France, 1977. *120 Years of Electronic Music*. https://120years.net/upic-system-iannis-xenakis-france-1977/

Di Scipio, A. (2023). *"UPIC." In A Xenakis Dictionary, edited by Dimitris Exarchos. Association Les Amis de Xenakis*. https://www.iannis-xenakis.org/en/dictionary-upic

Dow, R. (2023, February). *U&I Software MetaSynth & Xx CTX*. https://www.soundonsound.com/reviews/ui-software-metasynth-xx-ctx

Guliyev, E. (2022). *Violet Cold | creating Music | Patreon*. https://www.patreon.com/violetcold

John Cage | American Composer & Avant-Garde Innovator | Britannica. (2024, May 24). https://www.britannica.com/biography/John-Cage

John Cage Complete Works. (n.d.). Retrieved June 20, 2024, from https://www.johncage.org/pp/John-Cage-Work-Detail.cfm?work_ID=79

Lichtton Orgel (1935) | Digital Music Academy. (n.d.). Retrieved June 16, 2024, from https://www.digitalmusicacademy.ru/lesson-lichtton-orgel

Matthew, T. (2021, March 8). Space Oddity: The weird history of the Soviet ANS synthesizer. *5 Magazine*. https://5mag.net/features/history-soviet-ans-synthesizer/

Meyer, P. (2023, June 28). *Philip-meyer-max-tutorials/patchers/fundamentals-of-sequencing at main · pdmeyer/philip-meyer-max-tutorials · GitHub*. https://github.com/pdmeyer/philip-meyer-max-tutorials/tree/main/patchers/fundamentals-of-sequencing

Monastra, P. (n.d.). *Krzysztof Penderecki's Polymorphia and Fluorescences | A Guide to the Moldenhauer Archives | Articles and Essays | The Moldenhauer Archives—The Rosaleen Moldenhauer Memorial | Digital Collections | Library of Congress* [Web page]. Library of Congress, Washington, D.C. 20540 USA. Retrieved June 20, 2024, from https://www.loc.gov/collections/moldenhauer-archives/articles-and-essays/guide-to-archives/polymorphia-and-fluorescences/

Nechita, T. (2018). *MIDImage Download: Easy-to-use software application that allows you load various graphic images and convert them into MIDI audio files in no time at all.* https://www.softpedia.com/get/Multimedia/Audio/Other-AUDIO-Tools/MIDImage.shtml

Op de Coul, M. (n.d.). *Scala Home Page*. Retrieved June 17, 2024, from https://www.huygens-fokker.org/scala/

Optophonic Piano—Monoskop. (n.d.). Retrieved June 16, 2024, from https://monoskop.org/Optophonic_Piano

Recherches et inventions | 1934-10-01 | Gallica. (n.d.). Retrieved June 16, 2024, from https://gallica.bnf.fr/ark:/12148/bpt6k5830458b/f14.item

Rehabilitation industry of Russia. (2019). *Пионеры новых технологий: Вадим Арцев рассказал, как перестал быть незрячим / Хабр*. https://habr.com/ru/companies/russian_rehab_industry/articles/458532/

Sandhana, L. (2003). *BBC NEWS | Science/Nature | Blind "see with sound."* http://news.bbc.co.uk/2/hi/science/nature/3171226.stm

Sauerwald, M. (2024, June 20). *Media Art Net | Cage, John: Fontana Mix* [Text]. Media Art Net. http://www.medienkunstnetz.de/works/fontana-mix/

Scale Workshop · xenharmonic-devs/scale-workshop Wiki · GitHub. (n.d.). Retrieved June 17, 2024, from https://github.com/xenharmonic-devs/scale-workshop/wiki#overview

The ANS Synthesiser pt2.0 ANS Discography. (2007, November 22). *Stalker*. https://crab.wordpress.com/2007/11/22/the-ans-synthesiser-pt20-ans-discography/

The Dream of Color Music, And Machines That Made it Possible. (n.d.). Retrieved June 16, 2024, from https://www.awn.com/mag/issue2.1/articles/moritz2.1.html

The 'Optophonic Piano', Vladimir Rossiné, Russia and France. 1916 – 120 Years of Electronic Music. (n.d.). Retrieved June 16, 2024, from https://120years.net/the-optophonic-pianovladimir-rossinesoviet-union1916/

The 'Variophone' Yevgeny Sholpo. Russia, 1932. (2013, September 23). *120 Years of Electronic Music*. https://120years.net/the-variophoneyevgeny-sholposoviet-union1932/

the vOICe—The Index Project. (n.d.). Retrieved January 15, 2024, from https://theindexproject.org/post/the-voice

U&I software MetaSynth—Specifications, pictures, prices, links, reviews and ratings. (n.d.). Retrieved June 20, 2024, from https://sonicstate.com/synth/u_metasynth/

Warmplace Virtual ANS. (n.d.). Retrieved June 16, 2024, from https://www.soundonsound.com/reviews/warmplace-virtual-ans

WarmPlace.ru. Virtual ANS Spectral Synthesizer. (n.d.). Retrieved June 16, 2024, from https://www.warmplace.ru/soft/ans/#about

Wojcik, D. (2023). *Note by Daniel Wojcik on Substack: "Caveat: I can't find my original research, so this is a re-creation using an old brain with faulty memory. I can remember phone numbers from 1972, but not what I did last Tuesday. What's up with that? Anyway, way back during the last millennium, I began using MIDI to create …."* https://substack.com/@danielwojcik/note/c-39844230