SOUND CANVAS
DESIGNING AND EVALUATING EXPLORABLE SONIC LANDSCAPES USING EYE-TRACKING

Martin Linder Nilsson
KTH, Royal Institute of Technology
The Department of Media Technology
and Interaction Design
hmni@kth.se

Johannes Loor
KTH, Royal Institute of Technology
The Department of Media Technology
and Interaction Design
loor@kth.se

ABSTRACT

When we look at a piece of art, especially if it is non-abstract work, we explore the motifs and navigate through the surroundings using our eyes. Hopefully, if we find the artwork appealing enough, this instills an emotional response and creates in us a sense of immersion. Through the construction and evaluation of Sound Canvas, an audio-visual installation, this paper explores whether these or similar sensations as well as a sense of immersion can be attained without the usual visual feedback of a traditional painting - instead using another output modality, namely sound. Sound Canvas is a blank canvas in a picture frame housing 22 different sounds that observers can freely explore using eye-tracking technology and a pair of headphones. Through this project we investigate our hearing from an artistic view and entertain the notion that our hearing might be as good at navigating through and experiencing an artwork as our eyes are. Results from user testing show that navigation is indeed possible using only auditory stimuli. A sense of immersion was also attained by the participants, albeit in varying amounts.

1. INTRODUCTION

We humans have continually sought new ways of generating sound through history by constructing instruments and using different parts our bodies such as our hands, fingers, lungs etc. Advances in technology have in recent years opened up for a variety of new modalities and techniques for producing and controlling sound previously not available to us. One of these is the gaze modality, which up until recently hasn't been associated with sound and even now sees very little use in the field. This may be due to that the gaze modality can be problematic in controlling [1, 2] and producing sound. For our purposes however, eye-tracking technology provides the unique ability to use the gaze modality as an input while using sound as an output, which is the main focus of this study. Although sparsely used in sound applications, eye-tracking has seen a fair amount of use in other fields [3, 4], healthcare and gaming being amongst the most prominent.

To be able to investigate the viability of sound art, an installation was built. The installation is called Sound Canvas and consists of a frame with a blank canvas. The observer's gaze is tracked using eye-tracking technology and triggers the system to playback different sounds depending on where the user's gaze is located. Using the gaze modality as an input for this project lets us keep the same visual input modality of looking at a piece of art, only changing the mode of feedback from visual to auditory. In other words, the blank canvas lets us test whether the connection between using your eyes as input and generating sounds as feedback is prevalent even without the visual feedback component present in visual art. 22 different sounds were placed at various points on the canvas with the ambition to create an environment with three distinct characteristics: a nature theme, a city theme and an indoor theme. To produce a sense of navigation and to be able to locate the sounds, volume and panning is adjusted according to the gaze point's proximity to the sounds' center-points.

1.1 Related Work

Tracking the movement of the eyes for interaction purposes can be somewhat tricky for a few reasons. One being that even if you keep your gaze fixed at a singular point, the eyes still flutter a tiny bit which is unnoticeable for the person looking but is prone to create errors when using eye-tracking technologies. A study by Boyer et. al [5] showed that auditory feedback improved the awareness of, and ability to control, oculomotor behavior. By comparing Boyer et. al's system with Sound Canvas similarities are evident in how both systems generate sound feedback depending on the users movement, Boyer et. al calls this "movement sonification" by referencing [6]. Even if learning to better control one's gaze through movement sonification is not the main objective of the present study, it is a possible side effect and could be an interesting topic for future works.

Connecting eye-tracking to musical expression is something that has been explored in previous research, mostly by making different kinds of gaze-controlled instruments. One example being a project by Vamvakousis & Ramirez called "EyeHarp" [2] that enables people with motor impairments to create music through interaction via gaze. Using EyeHarp the user controls a step sequencer to build up chords and arpeggios whilst playing a melody by looking at representations of notes. The results show the EyeHarp to be, like most traditional instruments, difficult to master but at the same time being a platform where the user can express their musicality at different complexity levels. These results were carefully considered when designing the Sound Canvas.

A pioneer in the field of eye-tracking and sound is Andrea Polli. In her paper "Active Vision: Controlling Sound with Eye Movements" [7] several of her projects in this area are presented together with insights and conclusions. One of these projects, called "Inside the Mask", was a collaboration between Polli and vocalist Carol Genetti where Polli built two systems, one for each performer. Genetti was able to add harmonies to her own vocals via a system that tracked her shadow, while Polli controlled the playback of recorded sounds by using an eye-tracking system. The latter system had about 30 recordings of voices in different languages laid out in a grid of squares on a screen. Polli was then able to control precisely which text to play as well as the amplitude of the sounds, using only gaze as input. An interesting insight by Polli, in regards to the context of the current study, was that she as a performer learned to control the system very precisely. This is always a problem area when working with eye-tracking as we are not used to using our eyes as input devices in the same manner as we would our hands. Regarding the graphic interface of the eye-tracking system, Polli comes to this conclusion: "In the design of the interface for eye-tracking music, it seemed that the simpler the visual aspects of the interface, the more effective it was for playing music." How this result compares to having no graphic objects, as is the case with Sound Canvas, is of interest for this study, even if it might not be directly applicable as Sound Canvas is not a musical instrument per se.

Another example of gaze controlled instruments is the SynthesEyeser, a project co-produced and built by the authors [1] . The SynthesEyeser consists of the same Tobii Eye-Tracking hardware used in the present study, letting the user control the cursor of a piano-like custom built interface and thus play melodies. The system varies the melodies pitch following the cursor along the x-axis and filters the sound along the y-axis. As opposed to the Sound Canvas, the interface has visual cues on were to look to achieve the desired note or filter amount. Design wise, this made sense when working with an instrument but also proved challenging as the interfaced needed to match the users expectation of how to use the instrument precisely. As the present study did not need clear visual cues to oblige any sort of music related set of rules, many challenges faced when building SynthesEyeser was eliminated by purposely keeping the canvas itself completely blank. This was intended to help the user focus on the sounds, and thus making the sonic experience more immersive. The lack of visual elements in interactive displays might go against ones instinct at first, since we are so used to having a visual representation guiding our gaze. However, it does not necessarily affect how effective the user is at completing tasks via the system. A study by Møllenbach, Hansen & Lillholm [8] showed this by letting participants use a gaze-controlled system to complete tasks, such as moving between fixed points on a screen, both with and without help from graphic display objects. The results showed no significant difference between the two settings, regarding both selection error rate and task completion time.

[1] The paper regarding the design and testing of SynthesEyeser is not yet published. Please contact the authors for access.

Sound canvas should however not be seen as a musical instrument, nor does it act as one, but instead it belongs in the field of art installations. When looking at art projects, gaze and eye-tracking has not been as widely implemented as sound installations. A sound installation that bares many similarities to the current study, working as an inspiration, is Random Access Lattice by Gerhard Eckel [9]. The installation lets the audience explore various sound recordings, each pinned to a specific location in a 2x2x2 m space, by moving a handheld speaker through the space. The speaker has sensors attached to it so its exact position can be tracked and then play location specific sounds, human voices in different languages, that vary in speed (not pitch) in accordance to how fast the speaker is moved. Though the main notion of pinning explorable sound recordings within a specific area is shared between Eckel's piece and Sound Canvas, there are some variations mostly connected to the technologies used. The main of which is the use of 2d-space in Sound Canvas instead of 3d and the control of panning and volume as opposed to speed. These differences will be more evident and the details more clear after reading the method section below.

1.2 Aim and Hypothesis

The aim of this project is to create a gaze controlled sound installation and to investigate whether sound could be used as an alternative to visual feedback when observing an artwork. If this is the case, is it possible to explore a landscape using this technology? And how does the sense of immersion differ from that of its visual counterpart?

Our hypothesis going in to the project is that this type of sound art should be able to instill emotions in the observer and create a sense of immersion. Even though sound and music is an extremely powerful tool in conveying emotion, it however might prove hard to create a sense of navigating through a landscape. This may largely be due to the novelty of the eye-tracking technology. Gaze tracking does not see much use in many applications, thus, few people have experience using the technology.

2. METHOD

2.1 Sound Canvas

Sound Canvas is an sonic art-installation consisting of a classic painting frame containing a blank canvas, an eye-tracker connected to a computer running the Pure Data software and a pair of headphones. The computer screen is "mirrored" to the canvas by placing the eye-tracker beneath the canvas instead of the computer screen, thereby providing a replica of the screen, hence letting the gaze feedback on the canvas register on the computer. The 22 prerecorded sounds were placed at various points on the canvas, all with different radii and position. This "placement" was done using Pure Data patches which pins the sounds to a pixel on the computer display according to a chosen x- and y-axis position (see figure 1). Furthermore, the patch uses the radius, R, determine the "size" of the sounds, i.e how large part of the screen the sound occupies. In effect, this creates circularly shaped sounds, of various sizes, placed at multiple points on the canvas. The observer's gaze point is tracked and its location determines which sounds are played at any given time, which allows the observer to explore different sounds and environments in different parts of the Sound Canvas.

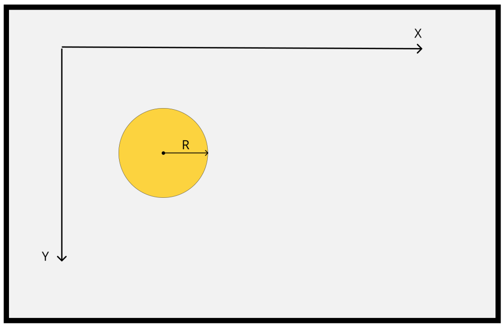

Figure 1. Image displaying how the sound circles were placed on the screen according to their x, y and R-value.

Pure Data logic also calculates the panning and volume of the sounds according to the gaze point's proximity to the sounds' position. The volume calculation returns 100% volume at the sound's center-point, with a linear decrease ending at 0% at the sound's R-value from the center-point. In effect, this makes the previously mentioned sound circles have the highest sound intensity in the center of the circle, gradually decreasing as you get closer to the edges (see figure 2). The panning calculation works in a similar fashion returning 100% volume in the right channel and 0% in the left channel at the R-value to the left of the sound's center-point and vice versa, resulting in 50% volume in each channel in the sound's center point. Combining the volume and the panning calculations nets in a feeling of distance to the sounds and were created in an attempt to

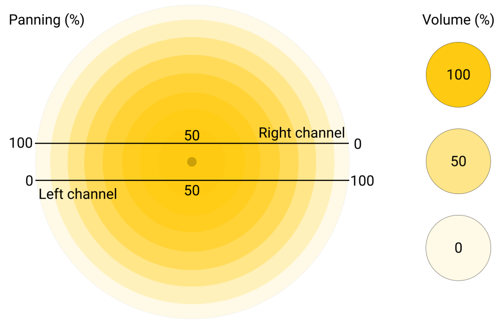

Figure 2. Image showing how the volume and panning of each individual sound was handled in Sound Canvas.

produce a natural sense of travelling through audio environments as well as to give the user the ability to navigate, find and pinpoint sounds using their sense of directional hearing as is done in everyday life. This method aims to place ”you” at the gaze point, hearing sounds coming from both directions. Moving your gaze, or ”you”, closer to any sound naturally makes them louder and more centered.

The audio used for the installation was recorded using a Zoom H4n hand field recorder and its built in microphones at various locations in Stockholm, Sweden. To provide breadth and an interesting auditory experience, a decision was made to divide the canvas into three, partially overlapping, differently themed environments (see figure 3). Before starting the collecting of sounds, a careful thought process focused on what we wanted to do and which emotions we wanted to instill in the different areas of the artwork was conducted. This process resulted in a number of expressions or emotions for each of the sections. The left-most third of the canvas, informally called *the traffic theme*, was thought to represent a busy city, featuring a rather noisy collection of sounds ranging from heavily trafficked roads and a tram leaving the station to bus sounds and car horns. In the lower half of the remaining area sounds from an indoor setting is found. For this *indoor section*, our thought process had rendered the expressions ”close by” and ”hushed” and therefore, a lot of the samples was recorded in a library. Altogether, the audio of this section provides a stark contrast to the traffic theme by using sounds like people whispering, a printer working, turning pages in books and the slight squeak of a chair. The upper right part of the Sound Canvas features *the nature theme* in which the aim was to create a peaceful environment, adhering to the expressions ”open”, ”nature” and ”serenity”.

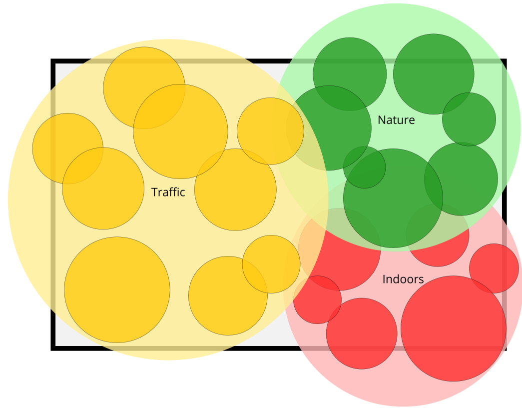

Figure 3. Image depicting the layout of the individual sounds categorized by color, yellow being traffic-themed, green being nature-themed and red being indoor-themed. The less saturated circles represent the over-arching theme sounds of each category.

To accommodate this, sounds like birds chirping and dogs playing was used.

To create the illusion of travelling through three distinct environments, each of the sections had one over-arching thematic sound spanning the entirety of the section. For the traffic theme for example, this was the sound of a city square featuring the audio of a typical city such as people talking and distant traffic. On top of this theme, more distinct sounds with smaller radii was placed, such as a bus stopping, creating the illusion of exploring different parts of an area. The thematic sound of the indoor theme was the sound of a library creating a faint but still very present sonic signature, and for the nature theme the sound of a forest was used.

To further enhance this feeling of motion, a few transition-sounds was used to emphasize the movement from one area to another. For example, between the traffic theme and the indoor theme the sound of sliding doors opening and closing was placed. This technique was also used within sections, an example being the sound of a subway train leaving a station next to the sound of an escalator and the familiar beeping sound (at least to residents of Stockholm) of the tube latch.

2.2 Evaluation

Evaluating the Sound canvas was a process that arced over a large part of the design process. User tests were made at several points in the development of the installation and

Figure 4. Showing the setting during user testing with the canvas, eye-tracker and a user wearing headphones.

the results guided our work going forward. In other words, the finished installation looks and works quite a ways from our initial ideas. These pre-tests mainly came in to play when deciding between two or several design choices, like if to use visual stimuli as well as the auditory feedback or just the blank canvas.

The main evaluation of the Sound Canvas was done with the help of 8 test participants without prior knowledge of the project. Each participant got a brief explanation that they were going to experience a sound installation. He or she was instructed to sit in a chair, put the provided headphones on and to freely explore the canvas for one minute (See figure 4). After this, the participant got to answer the following survey question:

1. *What was your overall impression of Sound Canvas?*

The participant was asked to once again put the headphones on and explore the canvas. However, this time he or she was instructed to try to locate as many sounds and environments as possible. Following this two-minute session, the participants was asked to evaluate their experience with the following questions:

2. *What types of sounds/environments did you hear or experience?*

3. *How was your experience navigating through the artwork and pinpointing individual sounds? Was it intuitive, hard, confusing etc ?*

4. *Did you experience a sense of immersion while using the installation? if so, how?*

5. *Did you find it hard or problematic to use the installation? if so, why?*

3. RESULTS

The first question regarding how people experienced the installation was asked to pick up on people's first impressions of the artwork, with them having no prior knowledge of what to expect. This was to investigate whether the connection between the modalities used would prove intuitive when confronted with them without, or with limited, instructions. There were mixed results. Some of the participants seemed to get the hang of what the installation was and how to use it rather quickly while a few didn't quite understand the connection between where they were looking and what they were hearing. This also reflected in their survey answers, with two test subjects commenting that they were confused and didn't know what to do. However, almost all participants expressed that they thought the installation was interesting, using words like ”cool” and ”fun idea”.

During the second session however, when people had gotten instructions to try to find as many sounds as possible, the understanding of how to navigate through the installation was much higher amongst the participants. This was noticeable both in observing them interact with the installation as well as in the survey answers. The question regarding what different kinds of sounds and environments they had been able to locate showed that they, on average, were able to account for roughly one third of the 22 sounds used. This is impressive since, even though a few of the sounds might be considered to be rather distinct, like the sound of a subway latch, many of the sounds are of a rather ambient nature and may be hard to place, like the sound of a forest or traffic sounds.

The results of question 3 revealed that the participants did not feel comfortable using their eyes to control the sound. Almost all test subjects claimed that it was difficult and a few expressed comments like ”I'm not used to using my eyes in this way”. Many participants did however comment on the use of volume and panning. One person said ”At one time i heard a faint sound coming from the left, so I went there and found it”. Although the test subjects found the novelty of using the gaze modality to be a problem, many of them felt that they got better at using it over time, even with the test time being only three minutes.

The question regarding immersion yielded interesting results. Almost all test subjects experienced some kind of immersion, but at varying degrees. Common comments from the participants who claimed they felt a great deal of immersion had to do with how the panning and volume was handled, though not in as technical terms. They expressed experiencing spaciousness and a sensation of movement and exploration, using words like "surrounded by sound" and "travelling through landscapes". Others, however, did not feel the same degree of immersion. One participant did not feel immersed at all and a few others expressed they felt little to no immersion, in large part due to that they didn't have anything to look at. A few participants also pointed out that it would be easier to feel immersed with eyes closed, this would however of course render the installation unusable.

In the answers to the last question, regarding any problems they had using the installation or if they found it hard to use, the participants repeated the points expressed in previous questions. The most commonly stated problem was the one about the difficulty of using the gaze tracking technology followed by that people found it hard to focus on a blank surface.

4. DISCUSSION

Results show that using your gaze to navigate through a sound-environment without visual feedback is indeed possible, at least for most of the participants in this study. Following participants' claims of feeling themselves travelling through sonic landscapes, it is also safe to say that some degree of immersion can be attained, albeit in greatly varying amounts.

One of the main obstacles of this project is that people in general are not used to used to using their eyes as an input modality, i.e. as a method of controlling things. The use of this novel modality is however the thing that make this project interesting. To remove the feedback component from vision while keeping the input lets us investigate the link between vision and hearing in an interesting way. To separate vision input and feedback is however not an easy feat since they, of course, are inherently intertwined but we believe this method proved to be a valid attempt. The choice to keep the canvas blank was made to promote exactly this, to put a focus on the auditory output and to reduce other distractions. As other studies have shown [1,2], use of the gaze modality is hard to master and might not be optimal when it comes to controlling sound. However, that people state that they only after a few minutes of use, felt themselves improving in their control of the modality testifies that the biggest obstacle might not be the modality itself but the novelty of the technology. As talked about in the related works section, this coincides with Polli's [7] ability to very precisely control a eye-tracking system after hours of training and performing. Polli also discussed how the system became more effective as the graphics got simpler, which would be interesting to compare with the immersion rate in the present study. Even if a comparison between using graphical display objects and leaving the canvas blank was not conducted in the scope of this study, some state of immersion was displayed from almost all participants. This gives an idea of how important sound, and what we hear, is for our overall impression of a piece of art, even if it is framed as a traditional painting.

As mentioned previously, this project went through several iterations in the design process and with this also multiple crossroads in design choices before the final prototype. One of the points where the project might have taken another path is when it comes to how to handle the spacial aspects. The way we wanted the interaction to feel was that the gaze point is where "you" are located. With the method we used, moving your gaze means that the panning changes according to your proximity to the sound, creating a sense that you yourself move. There were however other options that were discussed during the design process. One was to pan the sounds according to their placement on the canvas, meaning a static placement panning-wise, with the sounds placed on the left side of the canvas being panned to the left and so on. This would have produced a totally different installation, but in our opinion not have produced the same degree of immersion as was sought for this project. It would however be an interesting endeavour for another project.

As the eyetracker used, Tobii model 4C, and its software has a screen size limit of 27" we had to come up with a way of making it support the notably larger canvas. Various methods were tried, for example placing the eye-tracker a distance away from the canvas and thereby creating a larger virtual screen. However, experiments showed that the most accuracy and satisfactory result was attained by changing to a smaller screen resolution in the computer settings.

The fact that a canvas was used to "project" the computer monitor's image on meant that we couldn't calibrate the eye-tracker for each participant. It is safe to assume that this led to slightly different experiences for the participants due to factors such as difference in height and distance between the eyes. To not be able to calibrate the eye-tracker might have led to it not responding as well as it should to user input and/or a slight offset to the gaze point. To mitigate participants' height difference a chair with adjustable height was used, assuring that both height and distance from the canvas was roughly the same for all test subjects. However, due to the difference in distance between eyes, the experience might have been altered for some participants.

4.1 Project assessment

The construction and evaluation of Sound Canvas was an interesting experience. The rather ambitious idea of trying to separate vision input and output started out as something quite abstract but became more and more concrete as the project progressed. The calculations and programming in Pure Data worked smoothly and was a perfect fit for our project, largely due to it being efficient. This fact let us work with a large amount of audio files with volume and panning being calculated in real-time. Work with the eye-tracking also progressed rather well. Both of us had previous experience of eye-tracking technologies and had even worked with the specified Tobii model, which saved us a lot of time and frustration getting to know the limits of the hard- and software. Unfortunately the Covid-19 pandemic delayed the work by a fair amount. It also made our pool of test subjects a bit smaller as a few of the would-be participants cancelled their appointments.

Through the course of this project we have learned both technical and design related things. On the technical side, a deeper understanding of Pure-Data was gained through learning to implement somewhat complicated calculations and getting data from the mouse-cursor position. A large part of the project was spent on designing this abstract piece of technology. Designing projects within the domain of art and being none visual is something new for both authors, and ended up in many lessons learned.

We, as co-designers and authors of this paper are used to and comfortable working together, meaning that we know each-others' strengths and weaknesses making it quite painless dividing up the work. That being said, we continuously made conscious efforts to always keep the other party in-the-loop so that we gave each-other the chance to be a part of every aspect of the project. Furthermore we find this kind of work to be truly fascinating and are both eager to learn which was motivating.

5. REFERENCES

[1] J. San Agustin, J. C. Mateo, J. P. Hansen, and A. Villanueva, "Evaluation of the potential of gaze input for game interaction." *PsychNology Journal*, vol. 7, no. 2, pp. 213–236, 2009.

[2] Z. Vamvakousis and R. Ramirez, "The eyeharp: A gaze-controlled digital musical instrument," *Frontiers in Psychology*, vol. 7, p. 906, 2016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2016.00906

[3] J. D. Smith and T. N. Graham, "Use of eye movements for video game control," in *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, 2006, pp. 20–es.

[4] E. M. Kok and H. Jarodzka, "Before your very eyes: The value and limitations of eye tracking in medical education," *Medical education*, vol. 51, no. 1, pp. 114–122, 2017.

[5] E. O. Boyer, A. Portron, F. Bevilacqua, and J. Lorenceau, "Continuous auditory feedback of eye movements: An exploratory study toward improving oculomotor control," *Frontiers in Neuroscience*, vol. 11, p. 197, 2017. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2017.00197

[6] F. Bevilacqua, E. O. Boyer, J. Françoise, O. Houix, P. Susini, A. Roby-Brami, and S. Hanneton, "Sensori-motor learning with movement sonification: Perspectives from recent interdisciplinary studies," *Frontiers in Neuroscience*, vol. 10, p. 385, 2016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2016.00385

[7] A. Polli, "Active vision: Controlling sound with eye movements," *Leonardo (Oxford)*, vol. 32, no. 5, pp. 405–411, 1999.

[8] E. Møllenbach, J. P. Hansen, and M. Lillholm, "Eye movements in gaze interaction," *Journal of Eye Movement Research*, vol. 6, no. 2, May 2013. [Online]. Available: https://bop.unibe.ch/JEMR/article/view/2354

[9] G. Eckel, "Random access lattice," 2011, accessed: 2020-08-13. [Online]. Available: https://users.iem.at/eckel/art/art.html