SynthesEyeser

Constructing and evaluating a gaze-controlled digital instrument

Linda Cnattingius - lindacn@kth.se

Martin Linder Nilsson - hmni@kth.se

Johannes Loor - loor@kth.se

Carl Zettergren - carzet@kth.se

ABSTRACT

Technological advances continues to enable new ways of creating and controlling music. One of the newer modalities to be explored in this context is gaze. This paper examines whether gaze is a viable modality for multimodal musical production and if musical experience has an impact on learning instruments that use novel interaction modalities. In order to investigate this, an instrument prototype dubbed the "SynthesEyeser" was developed with consideration to previous research of gaze interfaces and gaze in music. A low-latency micro-controller, Bela, powers the instrument and utilizes a Tobii eye-tracker for gaze control as well as an ultrasonic sensor for gesture control. The instrument was evaluated in an experimental test setting through participants' self assessment and qualitative experience. Results indicate that gaze can be a viable modality for music production, but has some issues with the natural uncommoness of the type of eye-interaction required. The test conducted in this study also showed no correlation between the amount of previous musical experience one has and how fast instruments using novel modalities can be learned.

INTRODUCTION

The art of creating music dates back thousands of years and we humans have continually sought after new ways of constructing instruments and making sound using our hands, fingers, lungs and other parts of our bodies. In later years, digital technologies have opened up a plethora of possibilities to express ourselves using methods and modalities previously not related to music. One of these modalities is gaze. There may be reasons why gaze is not a go-to modality when dealing with sound, in the sense that it might not be cognitively natural for us to control musical aspects with our eyes or that the data may prove hard to filter. But this might also be something that we are able to train and/or develop tools to deal with.

Gaze between people is constantly used in music contexts subconsciously, in communication between musicians for example. Using gaze to directly generate or control sound is however not as common. One instance of gaze being used in a sort of meta-production of music, might be musical conductors. The musicians in the orchestra produce the actual sounds but they are directed to do so by the conductor. The orchestra is their instrument, so to speak. In this sense, gaze is a natural part of music and music-making. This study will evaluate the use of gaze in a more active sense.

To be able to investigate the viability of the gaze modality in controlling a musical instrument, a gaze-controlled digital instrument we decided to call the SynthesEyeser was developed. The instrument uses the gaze's horizontal position on a screen to produce sounds in different pitch and the vertical position to control the amount of modulation applied to the sound. The user is also able to control the volume and mute the instrument using their hand's proximity to a sensor. Hence, the input modalities used are gaze and gestures, with sound and visual feedback as output modalities. An experimental test was designed where 13 participants got to use and try to learn the SynthesEyeser and evaluate their experience. The participants were also subject to quantitative evaluation as they got to assess their ability in the beginning and the end of the test respectively. In this

way measuring their progress learning the instrument.

In this paper we evaluate the study on the SythesEyeser by first providing an explanation and discussion of previous works on the subject. This is followed by a detailed explanation of the instrument as well as technologies and tools used in the construction. We then describe how the test was designed, the data analysed and our results and findings.

RELATED WORK

In the preparations for this study, a state of the art search was conducted to explore the field of eye-tracking and music. However, very few instances of working instruments were found. The "EyeHarp" project conducted by Vamvakousis & Ramirez (2016) is by far the most intricate musical instrument, entirely controlled by gaze, that could found during the state of the art search. It is built for people with severe motor disabilities, letting them create and perform music only using their eyes. The instrument consists of a step-sequencer layer and a melodic layer, in which the first is where cords are constructed and the latter is where these chords and melodies are played. The results of the study show that both performers and audiences feel that the instrument is capable of producing expressive performances. Results also indicate that similarly to traditional instruments the EyeHarp has a steep learning curve, where playing in tempo with the eyes is one of the biggest challenges.

The study conducted by Boyer et.al. (2017) measures if auditory feedback has an effect on oculomotor control. They examine an interesting part of the physiological functions of the eye, and therefore gaze. Namely that eye movement is generally not smooth (unless following a moving object or auditory stimuli), and even while fixating or resting the gaze on something the eyes still have very small muscle movements that are unnoticeable in perception. These small movements can produce errors in gaze tracking. Of note is that the participants who, for the first half of the experiment, were subjected to audio feedback had better results in the non-audio feedback sessions as they had gained more awareness of their oculomotor control. Worries about inaccurate gaze might therefore be dissipated somewhat as the audio-modality output might subdue gaze-noise as it might make it clearer how the eyes are controlling the instrument. During the discussion in Boyer et.al. the possibility of using clearer and more harmonic tones as feedback in further research might give users something more recogniseable to better understand and train their eye-movements to, this is something that influenced the sound profile for the SynthesEyeser.

Poggi (2002) analyzes performances of conductors leading orchestras with the intent of creating a catalog of meaning for the multimodal communication that conducting entices. One of the modalities analyzed is that of gaze and its role in the overall communication between conductor and ensemble. By utilizing gaze direction the conductor can tell the musical ensemble, or parts of it, to play their instruments in a specific way. A look down can mean things such as "I am not ready", and the absence of gaze (the closing of the eyes) can mean various things, but the most important aspect seems to be the act of addressing certain instrument groups that the other aspects of the current multimodal interaction is directed at them. Gaze therefore exists to some extent as a modality for musical production, and might be able to be used in other scenarios.

Møllenbach, Hansen & Lillholm (2013) provides a taxonomy of gaze interaction. The taxonomy enables a common scientific language for discussing various implementations of interaction and visualization. The secondary part of their paper was a study that came to the conclusion that graphic display objects are not always necessary for successful interaction, which was considered in the interaction design of this project. This was ultimately discarded for a graphical user interface (UI) implementation as this fit the research question and method better. If no aspects of traditional instrument modalities were implemented (in this case visible notes) it might have been harder to understand how musically experienced people drew connections to their experience, this gave them a clue about how the instrument worked. It would also be hard to give the participants enough time to fully acquaint themselves with a pitch-range untethered to any visual UI within the limits of this project.

Mohan et.al. (2018) details a system developed to avoid the so called "Midas Gaze"-problem that might occur in systems where gaze is the main modality for selection. They describe this problem as when a user mistakenly selects or uses a part of the interface. They provide good information about how to avoid the issue in the implementation of gaze in this project. Their solution was to create a double confirmation system instead of traditionally used prolonged fixated gaze used for selecting interactables in a UI. In their user testing this seemed to give more reliable control to the user due to an increase in accuracy of interaction. Even though this projects' final implementation of gaze interaction with a UI differed from the one presented in Mohan et.al. (2018), it is necessary to be aware of issues such as "The Midas Gaze" problem.

Aim and Hypothesis

The aim of this study was to evaluate an interface to create and control music using gaze-tracking and gestures. By doing this we hoped to get a better understanding of the gaze modality and if it, on its own or combined with other modalities, might serve as a viable part of music creation. As user studies were conducted and data collected from both people with and without previous musical experience, we also meant to investigate the effect that previous musical experience has on approach and learning curve of this new musical modality. Our main research questions are:

- To what extent is gaze-tracking viable as a modality for controlling a musical interface?
- How do musicians vs novices perceive interactions with a musical system based on the use of the modality gaze?
- Is our design usable and how can it be improved?

Our hypothesis is that this suggested interface and use of modalities will prove to follow other instruments in regards to a quite steep learning curve. We however suspect that the use of the gaze modality will seem unintuitive at first, since it is not commonly used to control technology. As for the interface design prototype we believed that the main concept of the design would be well regarded by the test-participants. and that these modalities in this type of interaction should be explored further.

METHOD

The instrument

The SynthesEyeser is an experimental digital instrument. By tracking the eye movements of the user with a Tobii Eye Tracker 4C, the pitch of the sound and the

amount of filter applied to the sound is controlled depending on where in the visual interface the user direct their gaze (Tobii, 2020). Pitch is controlled on the x-axis of the interface and the filter strength is on the y-axis. The Tobii Eye Tracker 4C was chosen due to it being a robust system for eye-tracking and due to its wide availability. Aside from the standard Tobii software that accompanies their products the additional "Gaze Point"-program was used for mouse emulation. The choice to use the eye-tracking for mouse emulation, instead of using raw data, provided a smooth and stable way of combining the eye-tracking input with the rest of the system. Also the matter of potential legal issues concerning licensing and purpose of use of the Tobii API, was a factor in this decision. By using the mouse-emulation program the cursor data could be tracked in the javascript based visual interface and sent to the Pure Data patch.

Pure Data was chosen as it is a fast and versatile tool in terms of sound generating and processing. Aspects of modularity was also taken into consideration. An aim of the project was to make it possible to add new or change sounds in the system, in effect creating a completely new instrument. The Pure Data sound-patches can be added or exchanged quite easily to achieve this. In our current prototype of the instrument, two separate sounds with different types of modulation is implemented. The user can easily switch between these two audio profiles using a physical button. One of the two implemented profiles consists of a compressed sine wave sound with a granular delay as modulation and the other is a square wave sound modulated with a tremolo.

The volume is controlled by gestures, using an ultrasonic proximity sensor. Keeping your hand steady in front of the sensor causes the instrument to play a note of constant volume and removing the hand causes the sound to stop. Moving it closer or further away from the sensor lowers or increases the volume respectively, ranging from very low volume at 0 cm to maximum volume at 40 cm. An integral part of the design of the instrument was to be able to play constant fluid sounds as well as short notes in different tempi, providing versatility and the feel of a common instrument that affords physical touch interaction. Findings by Vamvakousis & Ramirez (2016) show that one of the biggest difficulties in playing their gaze-controlled instrument was to control the tempo with the eyes. We tried to cater to this by making it possible to play shorter notes in different length and tempi by rapidly blocking and unblocking the sensor with your hand in a "chopping" motion, as well as being able to play dynamically varying notes utilizing the proximity to the sensor.

The code for the sensor, the button, and the visual interface all run on a Bela microcontroller (Bela, 2020). Bela was chosen due to its low latency when dealing with sound and its compatibility with sensors. The Bela runs a Pure Data project with a wrapper written in c++ to handle the pulses coming from the ultrasonic proximity sensor. The visual interface is written in P5.js which is a JavaScript library integrated to work well on the Bela. The UI was designed to somewhat inhibit the effects of small eye-movements and sudden saccades that could produce a "Midas Gaze" issue by having a slight delay and averaging the cursor data before positioning of the blue ball cursor. This cursor also provided redundant visual feedback of pitch and modulation. Higher or lower pitch was visualised with a smaller respectively larger ball. More modulation made the ball turn a more

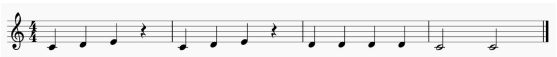
saturated blue color, while less modulation turned the ball into a desaturated blue. The SynthesEyeser is compatible to run on all screens up to 27", due to the limitations of the eye-tracker, and for our tests we used the maximum size of a 27" monitor to present the visual interface.

User tests

The participants were given general instructions of how the instrument worked, and what they were expected to do during the test on a piece of paper. Then they went through a calibration of the eye-tracking using the Tobii standard calibration software. This gave the participants a quick introduction to the eye-tracking modality. After a quick calibration the participants received control of the instrument. They were then prompted to test out the different functions of the instrument in four tasks:

1. Raise and lower volume, to acquaint them with the gesture based volume control.
2. Change pitch by shifting gaze across the horizontal axis.
3. Change modulation by shifting gaze across the vertical axis.
4. Switch the sound profile with the yellow button and then switch back to the first sound.

When these tasks were done the participants were prompted to play a classic swedish tune, 'Spanien', with the simple melody:



This song is commonly used in elementary school musical education in Sweden and it was chosen for this test due to its simplicity and recognizability. The participants tried to play this tune about 3 times before being prompted to fill in a ten point scale self-assessment form, answering the question "In your opinion, how well were you able to play the melody using the instrument?". An answer of one represented "Very poorly" and ten "Very well". Then the participants got five minutes of free-play time were they were prompted to learn the instrument to the best of their ability. They were not informed about how long this period was, only that they would be told when it was over. If they did not discover it on their own, they were told halfway through the time period that removing the hand muted the sound, providing an analogy to the notion of touch in other instruments (e.g. striking a key, strumming a string etc.). After the time was up, the participants were prompted to play the melody "Spanien" once more and fill in the same self assessment form as before. At the end of the test they filled in a form with free-text questions, used as the foundation for a qualitative evaluation.

RESULTS

The 13 participants were between 21-35 years old, three of them had used an eye tracker before and 11 had some sort of musical background. Various stroke and wind instruments were a part of some users musical background but the piano and the guitar were the most commonly played instruments, of which 10 played the former and 6 played the latter. The high number of people experienced with a keyboard instrument could have affected the general proficiency in learning to play the SynthesEyeser, because of the basic resemblance in note placement between the two. If the user pool had consisted of more people with no experience or knowledge of the fundamentals of the piano, the results may have been different. The amount of musical experience were more widespread than anticipated in the selection of participants. They were split into two groups for analysis. One group contained participants who were more

actively using musical instruments in their day to day life, while the other group was composed of people who had only occasionally dabbled with music, or had no extensive experience.

Quantitative Evaluation

The change in self assessment between the two "Spanien"-playings were compared between these two groupings using an ANCOVA test.

Little/no experience		Experienced	
Beginning of test/ concomitant variable	End of test/ dependent variable	Beginning of test/ concomitant variable	End of test/ dependent variable
5	5	7	7
5	6	4	7
7	6	7	9
7	9	7	7
4	6	4	7
		3	5
		5	8
		6	8

Assessment data used in the ANCOVA for 2 independent samples, p-value = 0.152378 > 0.05 indicating no significant difference between the sampled groups.

The resulting p-value of this ANCOVA test, 0.152378, was bigger than 0.05 which means that there were no significant difference in self-assessment change between these two groups with a 95% confidence interval. More experienced, or rather: "active", musicians did not get more comfortable any quicker than people who lack the same experience and activity. Looking at all participants as one group and comparing their first self-assessment with their second using an ANOVA test gives a p-value of 0.012885, and therefore indicates a significant change with a 95% confidence interval. This might seem trivial, that the participants felt more comfortable in their use of the instrument after some practice, but this shows that learning to control this implementation of the modalities is not completely impossible.

Qualitative Themes

In this part of the study we will present the qualitative results from the questions asked in the form, which the participants answered at the end of the experiment. This will be done by thematic analysis, presenting recurring themes in the answers to the form (Ryan & Bernard, 2000).

Impressions

When answering the question "What was your first impression?" several participants showed appreciation and excitement for using this new modality for creating music. When describing the experience many participants used words such as "cool" and "interesting". One participant described the experience rather eloquently as:

"...it felt like I was singing with my eyes"

However, a few participants expressed that the instrument was initially quite difficult to control and required a considerable amount of focus. This could possibly be due to the novelty of using gaze-tracking a main interaction modality for instruments.

After becoming more familiar with the instrument and given a second opportunity to play the melody the participants had quite divided opinions about whether the instrument became easier to use. While some of the participants thought it became more accurate after some time, others thought that the instrument was still quite awkward to play. Some also expressed difficulties with reaching the edges of the interface with their gaze, and others thought is was difficult to play tones that

were too far apart. The accuracy of the gaze is highly dependant on how well the Tobii was calibrated after the participants gaze, and could therefore have affected the opinions of the user experience.

When asked "How did you approach using the instrument?" many participants mentioned getting the coordination between hands, eyes and musical intent to work, as a first step in learning the instrument. Some participants tried to use their previous knowledge of how to play other instruments as a starting point. For example, one user tried to play a C-major scale without accidentally playing notes outside of the scale. One user approach the instrument like this:

*"*As if I am playing harp with my gaze, based on how the interface looks like*"*

Controlling the instrument

The feeling of being in control of the instrument varied between users. Some felt in control and were, within the time period of the test, able to use the instrument as they intended, others found it more difficult. Common reasons given for these difficulties included input lag, lack of consistent focus needed, notes being visually too close together, difficulties in being precise when controlling the "mouse"-cursor and volume control issues. For the most part, the input lag issues concerned the output volume and was present because of a short delay, caused by the sample rate of the sensor as well as a ramp function added in Pure Data. This function was needed so that rapid changes in volume would not result in an unpleasant and distracting clicking noise.

Regarding the visual interface the participants expressed an overall positive opinion, describing it as visualising the changes well and the color coding of the lines indicating the tones as convenient. However, several participants thought that the lines could have been thicker to make them more visible. The vertical placement of the tones (C, C#, D… etc) was also discussed, as some of the participants wanted to play in the middle of the interface but their gaze tend to pull towards the letters which are located at one third and two thirds of the screen's height respectively. This implies that the participants relied on the letters to play the notes they intended.

DISCUSSION

Gaze is indeed an uncommon modality for music production. Even though instances of it can be found in ensemble-contexts such as the orchestra in Poggi (2002), that context is only a proxy of audio production. A more direct control of the sound with gaze is difficult to find. The implementation of it in the SynthesEyeser instrument provided a novel experience for the participants, regardless of musical experience. One of the main things that the participants expressed as lacking was the sense of musical touch, or note on/off. The tangibility commonly found in almost all other instruments, strings, keys, drum skins etc. that creates sounds was hard to translate. This might have been different if our continuous double octave scale would have been a quantized scale instead. In such a scale moving the gaze would create a bigger change in sound. The method chosen for the "touch" of the SynthesEyeser was the interruption of the invisible ultrasonic cone. This might have been too abstract, although similar to a theremin, for the participants too grasp in the time given. These issues led to a sort of "midas-touch" with this gesture interface, where users accidentally disrupted the sound when the cone-shape was not obvious and when hand-shape, and placement, was not optimized. User unfriendliness aside, this abstraction was

somewhat intended and in line with the research question of how valid novel modalities are in musical production.

In our results we found no correlation between the amount of previous musical experience the participants had and the rate in which they learned the instrument. The cause of this might be attributed to the small amount of test subjects, meaning that a larger test group may have provided a different result. Another factor that might have influenced this is how the division into sub-groups was made. For the sake of comparability we wanted one group with extensive music experience and one group with no experience. Our test group was however not that binary, far from it in fact. In hindsight, when assembling participants for the study we preferably should have advertised for people who were either complete novices when it comes to music or people with extensive training.

An alternative modality for touch, instead of the ultrasonic proximity sensor used to control the volume, would be to use sound pressure in the form of blowing into a microphone. This might reduce the hand-to-eye coordination problem and possibly make the learning process faster. Using "blowing" for instruments is also very common, such as in all wind instruments, which could provide a more natural feeling using the instrument. Blowing was also something that was considered in Vamvakousis & Ramirez (2016) as an alternative for selection in their system for users with motor disabilities. The blowing that would be required in the interface of the SynthesEyeser would require longer blowing than short bursts for selection, but in the same way it would remove the issue with hand-gestures and motoric-skill.

Another interesting adjustment to the SynthesEyeser could be to alter the visual interface. As Møllenbach, Hansen & Lillholm (2013) mentions, having visual feedback is not a necessity for gaze-based instruments. At one point in the development process there were ideas about leaving out the screen and visual feedback entirely, making something not bound to use on a screen, but we figured that this would increase the difficulty of controlling the instrument dramatically. It would however have been interesting to test whether removing either the ball indicating where you are looking, the lines displaying where the notes are or the entire screen, to see how that would impact the user experience. This might have provided a more "free" interaction experience. As we previously discussed, this instrument is not optimal for playing specific notes and therefore removing the lines and tone indications could emphasize and provoke more abstract melodies and experimental use of the instrument.

The output modalities, sound and visual feedback, might be considered to provide redundant information as the visual feedback is a representation of the sound produced. The input modalities, gaze and gestures, work in a complementary fashion. One might even argue that the button used for changing the audio profiles would count as the instrument using haptics as an input modality as well. Even though this might seem like a stretch with the setup and the sounds in the current prototype, a different variation of the setup might invite for a more frequent and musical use of the button. During the development process several other features and more buttons were planned, such as being able to change octave up or down with two buttons. Another button-related feature was to be able to change the main sound and the modulation filters separately, letting you combine these in which ever manner you would like. Although possible, this idea had to be

scrapped as different filters had to be implemented in quite complicated ways and at different places in the signal chain, making this modular approach hard to actualize.

CONCLUSION

This paper studied whether gaze is a valid input modality to control an instrument, in combination with gestures as a secondary modality. Our user study showed that gaze is a promising modality and may very well find viable implementations in future digital instruments. Even though the instrument was perceived as complicated and hard to control, this might not differ substantially from other instruments. Further studies should be conducted on whether the learning curve follows that of traditional instruments. No difference in how well or fast participants learned the instrument, regardless of musical experience could be seen.

For future implementations the note on/off touch of the gaze-instrument must be further researched and some considerations should be taken regarding the range of pitch that should be playable.

REFERENCES

Boyer, E. O., Portron, A., Bevilacqua, F., & Lorenceau, J. (2017). Continuous Auditory Feedback of Eye Movements: An Exploratory Study toward Improving Oculomotor Control. *Frontiers in neuroscience*, 11, 197.

Mohan, P., Goh, W. B., Fu, C. W., & Yeung, S. K. (2018, October). DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 79-84). IEEE.

Møllenbach, E., Hansen, J. P., & Lillholm, M. (2013). Eye movements in gaze interaction. *Journal of Eye Movement Research*, 6(2), 1-15.

Poggi, I. (2002). The lexicon of the conductor's face. *ADVANCES IN CONSCIOUSNESS RESEARCH*, 35, 271-284.

Ryan, G. W., & Bernard, H. R. (2000). Techniques to identify themes in qualitative data. *Handbook of Qualitative Research. 2nd ed. Thousand Oaks, CA: Sage Publications*.

Vamvakousis, Z., & Ramirez, R. (2016). The EyeHarp: A gaze-controlled digital musical instrument. *Frontiers in psychology*, 7, 906.

Bela. (2020). Bela homepage. Retrieved January 22, 2020 from https://bela.io/

Tobii. (2020). Tobii eye-tracker 4C. Retrieved January 22, 2020 from https://gaming.tobii.com/tobii-eye-tracker-4c/

Appendix 1- Video summary of the project
https://vimeo.com/386309256

Appendix 2- Initial instructions and consent form.
Presented here as it looked for the participants.

The eye tracker registers where you look on the screen.

We will start with a calibration of the eye tracking.

Your task will be to evaluate the instrument. We want to know what was positive/negative.

To start with you will get a few tasks to get to know the instrument.

Then you will get to play a melody.

Then you will get some time to play around with the instrument.

This is how the instrument works:

Adjust the volume by moving your hand along the black line on the paper. The audio turns off if your hand is not in line with the bigger box.

Adjust pitch by moving your gaze horizontally.

Adjust modulation by moving your gaze vertically.

The yellow button switches sound. There are two sounds.

Get to know the instrument:

1. Raise volume to maximum, then lower it to the lowest level.
2. Try and play some different pitches by moving your gaze horizontally.
3. Try and modulate the sound by moving your gaze vertically while you keep the same key.
4. Try and switch the sound with the yellow button. Play a little with this sound and then switch back.

Appendix 2-Continued.

Take some time to memorize the following melody (“Spanien”).

C D E C D E D D D D C

Now play the melody you memorized. You can look back at this note if you forget, but try and keep focus on the screen while you play.

Now you will get some time to play freely to get to know the instrument better. Try to learn the instrument to the best of your abilities.

Now you are going to play the melody “Spanien” again.

C D E C D E D D D D C

Appendix 3- Evaluation questions.

Self assessment questions:

1. In your opinion, how well were you able to play the melody using the instrument?
2. In your opinion, how well were you able to play the melody using the instrument?

Qualitative form questions:

- Age
- How much musical experience do you have? How many years of experience playing an instrument, and/or musical education.
- Which instrument(s) do you play?
- Have you used eye tracking before?
- What was your first impression?
- How did you approach using the instrument?
- What was your impression at the end of the test?
- Did you feel that you were able to control the instrument as you liked?
- Why/why not?
- Did you have any difficulties when playing the SynthesEyezer?
- Any thoughts about the visual interface?
- Anything you would like to add?