A Gaze-Driven Digital Interface for Musical Expression Based on Real-time Physical Modelling Synthesis

Kandpal, Devansh ; Kantan, Prithvi Ravi; Serafin, Stefania

Published in:
Proceedings of the 19th Sound and Music Computing Conference, June 5-12th, 2022, Saint-Étienne (France)

DOI (link to publication from Publisher):
10.5281/zenodo.6798248

Creative Commons License
CC BY 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Kandpal, D., Kantan, P. R., & Serafin, S. (2022). A Gaze-Driven Digital Interface for Musical Expression Based on Real-time Physical Modelling Synthesis. In R. Michon, L. Pottier, & Y. Orlarey (Eds.), *Proceedings of the 19th Sound and Music Computing Conference, June 5-12th, 2022, Saint-Étienne (France): SMC/JIM/IFC 2022* (pp. 461-468). Sound and Music Computing Network. Proceedings of the Sound and Music Computing Conference https://doi.org/10.5281/zenodo.6798248

A Gaze-Driven Digital Interface for Musical Expression Based on Real-time Physical Modelling Synthesis

Devansh Kandpal
Aalborg University
dkandp20@student.aau.dk

Prithvi Ravi Kantan
Aalborg University
prka@create.aau.dk

Stefania Serafin
Aalborg University
sts@create.aau.dk

ABSTRACT

Individuals with severely limited physical function such as Amyotrophic Lateral Sclerosis (ALS) sufferers are unable to engage in conventional music-making activities, as their bodily capabilities are often limited to eye movements. The rise of modern eye-tracking cameras has led to the development of augmented digital interfaces that can allow these individuals to compose music using only their gaze. This paper presents a gaze-controlled digital interface for musical expression and performance using a real-time physical model of a xylophone. The interface was designed to work with a basic Tobii eye-tracker and a scalable, open-source framework was built using the JUCE programming environment. A usability evaluation was carried out with nine convenience-sampled participants. Whilst the interface was found to be a feasible means for gaze-driven music performance our qualitative results indicate that the utility of the interface can be enhanced by expanding the possibilities for expressive control over the physical model. Potential improvements include a more robust gaze calibration method, as well as a redesigned graphical interface with more expressive possibilities. Overall, we see this work as a step towards accessible and inclusive musical performance interfaces for those with major physical limitations.

1. INTRODUCTION

Expression is a central component of human behaviour, and music is one of the oldest and most fundamental forms of expression [1]. Both conventional and *digital musical interfaces (DMIs)* and instruments [2] contribute strongly to human expression and interaction. Most forms of musical expression require some form of bodily movement, which means that individuals suffering from debilitating movement disorders can be severely limited in this regard. Advances in modern music technology have made it possible to address this through DMIs, and in recent years, many DMIs have been developed for individuals with physical limitations [3]. These have been termed as *adaptive digital musical instruments (ADMIs)*.

In a recent systematic review [4], ADMIs were segregated into categories including *tangible, touchless, biocontrollers, wearable,* etc., with tangible and touchless (camera-based) interfaces accounting for the majority. Of the remainder, there were only two gaze-controlled interfaces. Given the incredible communicative power of human eye and its role in regulating human interaction and emotional connection (and the advent of inexpensive eye-tracking technology) [5], it is interesting that the potential of gaze as an expressive medium has not been explored to a greater degree. Eye-driven ADMIs could, for instance, be invaluable to individuals suffering from ALS [6], who lose movement in all their limbs and are left with a very limited palette of interactive gestures. Moreover, existing work has proved that it is feasible to control sequencers and melodies using gaze [7, 8] to produce expressive performances. A gaze-based MIDI controller has also been developed [9] for use with digital audio workstations such as Ableton Live.

We believe that much of the potential of gaze-based ADMIs remains untapped. For instance, existing systems have typically used pre-recorded samples or simple oscillators for sound generation. Modern computing hardware has made it possible to manipulate more realistic-sounding physical models in real-time, which have been applied in interactive sonification research [10]. Physical models are particularly interesting for gaze-controlled ADMIs simply due to the sheer amount of parametric control that can be exercised over their sonic properties by a multitude of gaze parameters [5] with minimal physical effort. Current gaze-based DMIs are yet to explore modalities for interaction with real-time physical models, and to the best of our knowledge, there are no existing technological frameworks that integrate robust eye-tracking with physical modelling synthesis. Such a framework would greatly facilitate interdisciplinary and iterative user-centered studies, an indispensable component of human-computer interaction research [11].

The goal of the present work was to build a scalable and robust prototype framework to serve as a tool for exploring the interactive potential of gaze-controlled physical models. Subsequent sections cover the state-of-the-art in eye tracking and gaze-based DMIs, followed by the implementation, and usability evaluation of our prototype.

2. RELATED RESEARCH

The phrase *eye tracking* typically refers to the process of localizing the instantaneous gaze of a user. Gaze data is classified into two categories, fixations and saccades [12]. Fixations are definite, fixed gazes of a user at a certain spot, and saccades are rapid movements between two points that may also be involuntary [12]. Dwell time, which essentially is a measure of how long the gaze of a user dwells on a particular point, is a well known parameter that is used to differentiate between fixations and saccades [12]. In terms of design, eye trackers may be head-mounted or remote, and their working mechanism may be video-based tracking, infrared pupil-corneal reflection tracking, or based on electrooculography [5].

Gaze-based ADMIs have used both head-mounted [9] and remote [7, 8] eye-tracking systems. Aside from requiring accurate and regular calibration [5], fixation detection and smoothing algorithms have been found to be necessary for consistent system behavior [7]. The interaction itself has typically taken place on a graphical interface, with the gaze point serving as a *cursor* that manipulates visual elements. Interface design is constrained by limitations in the spatial and temporal resolution of eye-tracking systems [5]. Some existing guidelines are that on-screen artifacts must not be very small in size, and should be spaced out adequately and evenly [7, 13]. Other known ADMI design considerations relate to latency, inconsistent spatial accuracy, and the aptly-named Midas touch problem [8] where all viewed objects are selected [5].

From an interaction perspective, gaze-based ADMIs have provided control over step sequencers [7, 8], MIDI parameters [9], and melody notes [7]. However, the lack of time- and space-accurate parallel control over GUI elements translates to increased interface complexity for achieving standard music production tasks. In the task of playing a timed melody, the manipulation of timed sequencers addresses the difficulty of navigating quickly and accurately to on-screen note triggers, but adds a large number of interface elements. Note triggers also need to be arranged in a different manner than traditional instruments (e.g, in a ring [7, 13]), potentially making them less intuitive. In the only documented rigorous evaluation of a gaze-based ADMI, EyeHarp users [7] experienced steep learning curves and comparable difficulty in mastering the ADMI to a traditional musical instrument. The amount of expressive potential in existing ADMIs is also limited due to the absence of expressive integrated synthesizers; only simple oscillators and MIDI note triggers have been used [7–9], and more complex algorithms have yet to be integrated and experimented with.

Overall, gaze-based ADMIs are still at an early stage, and considerable advances are necessary in order to fully unlock the expressive potential of human gaze. An interesting possibility is that of integrating online gaze tracking with real-time physical modelling synthesis, a technique with the potential to generate realistic sounds, albeit at a relatively high computational cost [14]. We see this as a hugely promising direction as it can help achieve a potentially vast span of compelling and expressive parameter mappings to link gaze and musical sound. At this point, it is unclear whether a robust and usable implementation can be achieved for real-time operation. Therefore, the primary aim of the present work was to build a scalable prototype to (A) evaluate the feasibility of combining gaze tracking and physical modelling synthesis in a real-time interface, and (B) serve as an open-source framework to facilitate the future development of gaze-based ADMIs.

3. DESIGN AND IMPLEMENTATION

We developed a system integrating eye tracking, a gaze-controlled graphical interface and physical modelling synthesis into one software application. This section details the interaction design and implementation specifics. We used the following tools during the development process:

1. **MATLAB:** This was used to prototype the physical model of a single xylophone bar and tune the subsequent bars.
2. **JUCE:** For our software application, we used the JUCE framework for the real-time audio processing and GUI classes it provided.
3. **Tobii X2-30 Eye Tracker:** This model is of the remote variety, allowing for unencumbered head movement. We chose this for its simple interfacing with a windows laptop using the Tobii Eye Tracker Manager software, which allows user-friendly calibration of the device. The software was interfaced with JUCE using the C-based Tobii SDK.

3.1 Interaction Overview

1. The interface is run on a laptop, thus the screen acts as the performance space.
2. The eye-tracker is placed on a table in front of a user within its optimal range of operation (around 30-35 cm away)
3. The user calibrates the eye tracker with their sitting position before each use, and looks at the software interface (see Fig. 2), where a small green square indicates the current position of the user's gaze.
4. The user then plays music by looking at the various musical keys (GUI buttons). Additional controls for playing various types of diatonic chords and varying musical dynamics are also provided as buttons, which are similarly activated through gaze.

3.2 Physical Model of a Xylophone

The xylophone comprises multiple wooden bars each corresponding to a different musical note, which are struck using a mallet to generate sound. On breaking down the challenge of physically modelling a xylophone, it is observed that a wooden bar is the basic component of a xylophone. Wooden bars can be physically modelled based on the equations of the ideal bar provided by Bilbao in [15]. Chaigne and Doutaut also present the physical description of a xylophone bar in [16]. On inspecting the structure of a xylophone closely, it can be ascertained that the mallet is the 'exciter', and the wooden bars are 'resonators' in a

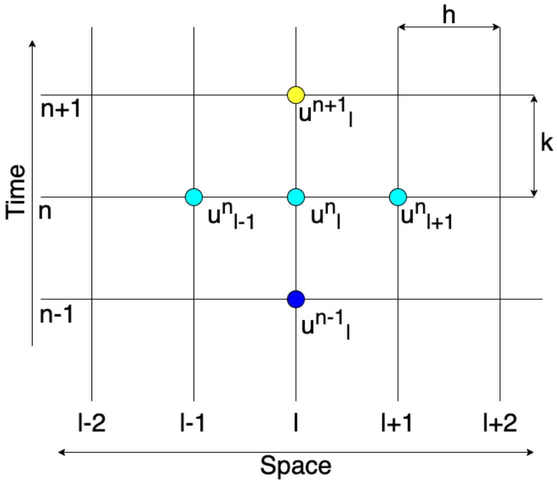


Figure 1. Graphical Representation of the Finite Difference Approach Used to Physically Model the Xylophone

xylophone. Vibrations are created when a mallet strikes a bar, which are sustained by the bar to generate an after-sound containing only the fundamental frequency of the bar [16] [17].

For our purposes, the sound of a xylophone bar was simulated by physically modelling an ideal bar using finite difference methods. To account for the decay of sound, frequency based and non-frequency based damping terms were incorporated in the equation of an ideal bar given in [15]. Further, the excitation action was simulated by using a raised cosine (Hanning window) as an initial excitation for the model. It was found that this method generated results that could be tuned to sound close to how an actual xylophone sounds.

3.2.1 *Mathematical Equations*

Bilbao provides in [15] the definition of the state of a system distributed in space as:

$$u = u(x, t) \tag{1}$$

The state equation of an ideal bar contains the second-order partial derivative of the state of the system in space and the fourth-order partial derivative of the state of the system in time, related by the stiffness coefficient of the ideal bar. It is given by Bilbao in [15] as:

$$u_{tt} = -\kappa^2 u_{xxxx} \tag{2}$$

Where u_{tt} denotes the second-order spatial derivative of u in time, u_{xxxx} denotes the fourth-order spatial derivative of u in space and κ denotes the stiffness coefficient of the bar.

The stiffness coefficient of the bar is calculated as:

$$\kappa = \sqrt{EI/\rho A} \tag{3}$$

Where,

- *E* denotes the Young's modulus of the material of the bar
- *I* denotes the moment of inertia of the bar
- *A* denotes the cross-sectional area of the bar, and
- ρ denotes the material density of the bar

The cross-sectional area of the bar is calculated by the product of its width, b and height, *x*. The following equation, obtained from [16] is used to calculate the moment of inertia of the ideal bar using the same two quantities:

$$I = bx^3/12 \tag{4}$$

However, equation (1) only describes an non damped, ideal bar. In real life applications, like the physical model of our xylophone, the bars are not be completely ideal and they contain damping factors that determine how quickly the generated sound fades away. The equation for a damped ideal bar contains two extra terms, one for frequency based damping and another for non-frequency based damping.

The non-frequency based damping term is given as:

$$-2\sigma_0 u_t, \tag{5}$$

The frequency based damping term is given as:

$$2\sigma_1 u_{txx} \tag{6}$$

Where σ_0 and σ_1 are the non-frequency based and frequency based damping coefficients, respectively.

Combined, equations (2), (5) and (6) give us the final equation that was used to model the bars of our xylophone:

$$u_{tt} = -\kappa^2 u_{xxxx} - 2\sigma_0 u_t + 2\sigma_1 u_{txx} \tag{7}$$

Equation (7) contains the following partial derivatives in addition to the ones in equation (2):

- u_t, which is the first order partial derivative of the state of the system in terms of time
- u_{txx}, which is the third order mixed partial derivative of the state of the system, in terms of both time and space

All the partial derivatives need to estimated using finite differencing methods in order to obtain the final equation for the state of the system. Furthermore, the stability of a finite differencing scheme is determined using a stability condition, that incorporates the spatial step of the discretised grid, *h* and also its temporal step, *k*. The temporal step, *k*, is calculated as the inverse of the sampling frequency.

The stability condition for a damped ideal bar, described by equation (7) is given as:

$$h = \sqrt{(4k(\sigma_1 + \sqrt{\sigma_1^2 + \kappa^2}))/2} \tag{8}$$

Finally, we arrive at the fully discretised state-update equation for an ideal bar, which is obtained by the expansion of all partial derivatives in equation (7) using finite differencing methods described in [15]. It is obtained as:

$\mathrm{u}_l^{n+1} = 2u_l^n - u_l^{n-1} - \kappa^2 k^2/h^4(u_l^n{}_{+2} - 4u_l^n{}_{+1} + 6u_l^n - 4u_l^n{}_{-1} + u_l^n{}_{-2}) - 2\sigma_0 k(u_l^n - u_l^{n-1}) + \sigma_1 k/h^2(u_l^n{}_{+1} - u_l^{n-1}{}_{+1} - 2u_l^n + 2u_l^{n-1} + u_l^n{}_{-1} - u_l^{n-1}{}_{-1})$

Here, u_l^n refers to the output of the system at the current time state (*n*) and at the current spatial stage (*l*). In this way, the locations of all terms in equation 4.9 can be ascertained.

3.3 GUI Elements and Layout

Twelve musical keys were provided, corresponding to an octave in an equitempered scale. Similarly to previous work [7], the musical keys of the xylophone were placed in a circular layout, with an inactive region in the center. The musical keys were coloured yellow, changing to green when struck to provide visual feedback. A green on-screen pointer was also incorporated to indicate instantaneous gaze position.

Finally, it was decided to not place GUI elements towards the extreme edges of the interface. This was done keeping in mind the fact that the accuracy of eye-trackers is higher towards the center of the screen, and it reduces closer to the edges of the screen.

3.4 Tuning the Physical Model

Once a working iteration of the physical model was obtained, it had to be tuned to sound like a xylophone. As a part of this process, the pitch and timbre of each key had to be tuned. The damping terms present in the state equation of an ideal bar were used to tune the sustain of the output. On the other hand, the length and height of the bar were altered to change the pitch of the model [17].

A collection of anechoic recordings of xylophone notes was used as a reference to tune the damping of the physical model by ear. These are made available for free by the University of Iowa [1]. The damping factors were altered until the output of the model seemed to have the same sustain as the anechoic recordings.

To arrive at the correct pitch for the xylophone bars of the ADMI, the MATLAB FFT function was used to analyse the output of multiple versions of the ideal bar. The next step was to tune each key to a fixed fundamental frequency. The fundamental frequencies of the semitones in the octave C4 - B4 are widely known, therefore the task at hand was to match the frequency of each of the 12 keys to that of the correct semitone. After some initial attempts, it was ascertained that increments in the height and the length of the ideal bar had the effect of increasing its fundamental frequency. It was noticed that an increase in the height of a virtual xylophone bar reduced the spatial resolution of the finite difference scheme, whereas an increase in the length had the opposite effect. Therefore, it was decided to only alter the length of the bar to bring about changes in the fundamental frequency. Bar lengths corresponding to one musical octave were thus ascertained, and can be seen in Table 1.

Note	Length (in metres)
C4	0.6810
C♯4	0.6720
D4	0.6520
D♯4	0.6460
E4	0.6270
F4	0.6191
F♯4	0.6030
G4	0.5690
G♯4	0.5356
A4	0.5270
A♯4	0.5120
B4	0.5061

Table 1. This table shows the different notes selected for the interface, and the length (in metres) of the virtual wooden bar for each fundamental frequency

3.5 Gaze Tracking Functionality

1. The Tobii X2-30 was interfaced with the computer using the Tobii EyeTracker Manager software on the computer [2].

2. The Tobii SDK [3] was used to interface the eye tracker with JUCE, which obtained instantaneous gaze data from the eye tracker at a rate of 40 Hz. The raw gaze tracking data was obtained separately for the left and right eyes.

3. A data description process was undertaken to permit fluid sonic interaction with the physical model. First, the two gaze points received at every instance were averaged to find a central gaze coordinate. Next, a gaze-detection algorithm was implemented to interpolate the data and allow smooth gaze based control. The gaze detection algorithm used was presented by Kumar et al. in [18], and its pseudo-code given by Kumar in [19]. This algorithm takes into account the previous coordinate at every step, and then compares it to a predefined 'saccade threshold' which is the radial distance from a point beyond which the subsequent coordinate is classified as a saccade. It then stores coordinates not described as saccades in a 20 sample window and computes the weighted mean of this window to obtain a final coordinate location. The on screen gaze pointer was programmed to follow this interpolated gaze coordinate location.

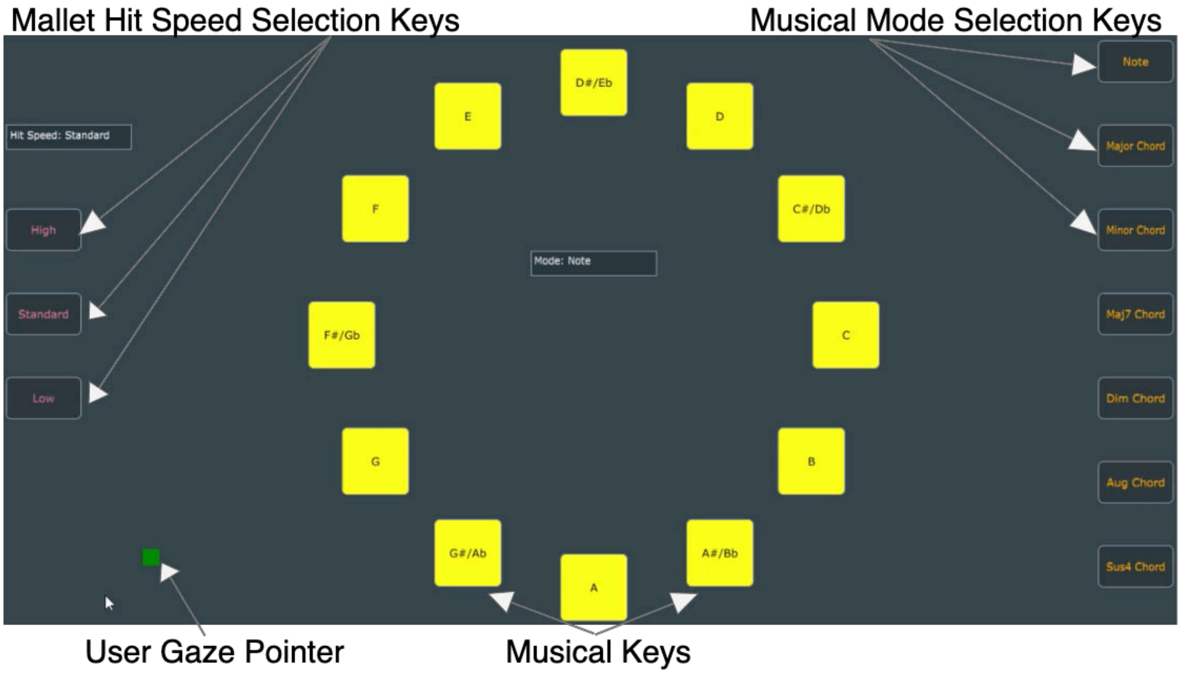
   The formula used to calculate weighted mean in the saccade detection algorithm is given in [19] as follows:

   p_{mean} = (1·p_0 + 2·$p1$ + ... + n·p_{n-1}) / (1 + 2 + ... + n)

[1] https://theremin.music.uiowa.edu/MIS-Pitches-2012/MISxylophone2012.html

[2] https://www.tobiipro.com/product-listing/eye-tracker-manager/

[3] https://www.tobiipro.com/product-listing/tobii-pro-sdk/#Downloads



Hit Speed: Standard

Mode: Note

Figure 2. User Interface of the Final Iteration of the Virtual Xylophone

4. Finally, a function was implemented to trigger any button that contained the on screen gaze pointer.

Audiovisual demonstrations of the xylophone and ADMI can be accessed on Dropbox [4]. The source code of the project is available on GitHub [5].

4. EVALUATION

We carried out a brief user study to assess the prototype in terms of

- **Usability** (captured using the *System Usability Scale (SUS)* [21])
- **Desirability and Playability** (using word selection techniques proposed in [22])

4.1 Participants

A convenience sample of nine participants (students at Aalborg University Copenhagen) was used for the evaluation. They had different levels of music training and general experience using eye-trackers. Each participant was informed about the purpose and duration of the experiment, and relevant consent was obtained from each participant. No sensitive personal information was collected about the participants and they had the option to withdraw at any time.

4.2 Experimental Setup

The experiment was carried out at the Multisensory Experience Lab at Aalborg University. It was hosted on an Asus laptop running the JUCE application with integrated eye tracking and also provided experiment instructions. Furthermore, a demo video of the interface was also shown on this laptop, which documented the tasks each participant had to perform as a part of the experiment.

4.3 Procedure

The following steps were followed by each participant:

1. First, the participant was briefed about how eye tracking works and the purpose of building this ADMI. The eye tracker was then calibrated.
2. The participant was then given some time to get accustomed to the ADMI and its interaction mechanism. Once a participant became comfortable with the ADMI, they were asked to play the C major scale on it.
3. The participant was then presented with a set of 3 different tasks, each of which was described to the participant through a video tutorial.
4. Finally, the participant was presented with the SUS survey and the desirability assessment word-list.

4.4 Data Analysis

The data were collected and analyzed in MS Excel. The key outcomes were as follows:

[4] https://www.dropbox.com/sh/rihsgakhpag0b6i/AACqLc5Eu-oiCa4pE-fXCWUIa?dl=0

[5] https://github.com/mrkandpal/EyeTrackingJUCE

Sr. No.	**Statement**	**Valence**	**Aggregated Score**
1	I think that I would like to use this system frequently.	+	2.34 ± 0.82
2	I found the system unnecessarily complex.	-	1.67 ± 0.67
3	I thought the system was easy to use.	+	3.12 ± 0.74
4	I think that I would need the support of a technical person to use this system.	-	1.67 ± 1.05
5	I found the various functions in the system to be well integrated.	+	4.00 ± 0.67
6	I thought there was too much inconsistency in the system.	-	2.34 ± 1.05
7	I would imagine that most people would learn to use this system very quickly.	+	3.78 ± 1.13
8	I found the system very cumbersome to use.	-	3.45 ± 0.95
9	I felt very confident using the system.	+	3.34 ± 0.94
10	I needed to learn a lot of things before I could get going with this system.	-	1.45 ± 0.50
11	**Overall SUS Score**		**65.5 ± 11.53**

Table 2. Results from the SUS questionnaire. Each statement was rated on a discrete scale from 1-5, where 5 indicated complete agreement, and 1 was complete disagreement. As is evident from the table, the statements alternated in terms of their valence (+/-). The aggregated values are presented as **mean ± standard deviation**. An SUS score of 68 is considered average [20]

1. SUS Score and its subcomponents

2. The occurrence frequencies of the selected words in the desirability assessment

5. RESULTS

Overall, we observed that the participants were able to use the interface and successfully perform the tasks. Table 3.5 summarizes the item-wise and overall aggregated SUS scores from the usability evaluation. The obtained mean **(65.5)** was slightly lower than the documented average from research [20]. The ratings indicate that that participants found the various functions of the interface to be well integrated **(Item 5)** and the overall system easy to use **(Items 2, 3, 4, 7, 10)**. Despite this, users appear to have found the system cumbersome **(Item 8)** and they did not express wishing to use it frequently **(Item 1)**.

Table 3 summarizes the outcomes of the desirability assessment. The majority of frequently- and somewhat used adjectives are positive in valence, with the exception of words such as *frustrating, clumsy, slow* and so on as shown in the table. The combined results are discussed in detail in the next section.

6. GENERAL DISCUSSION

In this work, we designed and developed a JUCE-based technical framework to explore the feasibility and interactive potential of gaze-controlled physical models for musical expression. We were successful in realizing a stable working prototype with simple note and chord generation capabilities, and our usability test showed that the graphical interface was operable without excessive difficulty.

Looking closely at the SUS results, it is clear that the participants did not find the system to be prohibitively complex or confusing. This could be attributed to the simplicity of the interface and its available interactions. The straightforward layout with clearly separated sections pertaining to hit velocity, note/chord mode and ring-shaped note triggers may have contributed to the participants' ease in operating the interface. The ring-shaped design was based on past approaches [7, 13] and appears to be a suitable layout shape for note triggers in such applications. However, it is equally likely that the simplicity of the interface contributed to participants not finding it engaging enough to wish to use it frequently. It will be interesting to see how participants' impressions of ease-of-use scale when the system itself scales in complexity and functionality. It is likely that the addition of new expressive affordances will lead to a similar steepening in learning curve to that seen in [7].

In spite of the stated ease-of-use, the overall usability of the system was rated below average, partially because participants tended to find the system cumbersome to use. We strongly suspect that this was in large part due to spatiotemporal inaccuracies in the eye tracking as well as calibration drift, both of which are well-documented issues with present eye tracking technology [5] and past ADMIs [7,9]. Whilst it is likely that newer hardware and firmware will gradually resolve these issues, a more sophisticated digital filtering algorithm (e.g, Kalman filter [23]) could have improved the usability of our system, and will be explored in future versions of the framework. Another possible cause of the user difficulty may have been the relatively brief training phase; it is likely that participants would have found the system easier to use with more practice, as also highlighted in [7].

The findings from the desirability evaluation aligned well with the SUS results, in that participants highlighted themes related to simplicity and intuitiveness, but also frustration and clumsiness. Whilst the overall positive disposition of participants can be interpreted in the system's favour, it is likely that being in the same room as the experimenter during the test may have introduced a social desirability bias into the responses of the participants. Future studies should adopt blinded designs with anonymous responses when conducting these assessments so as to minimize this

Frequently Used			**Somewhat Used**			**Seldom Used**		
Word	Valence	#Occ	*Word*	Valence	#Occ	*Word*	Valence	#Occ
Accessible	+	8	*Clumsy*	-	3	*Annoying*	-	2
Entertaining	+	5	*Easy-to-Use*	+	4	*Appealing*	+	2
Frustrating	-	6	*Responsive*	+	4	*Satisfying*	+	2
Intuitive	+	5	*Slow*	-	4	*Unpredictable*	-	2
Simplistic	+	6	*Stimulating*	+	3			
Straightforward	+	7	*Uncontrollable*	-	3			
			Unconventional	+/-	3			
			Usable	+	3			

Table 3. Results from the desirability assessment. The words are segregated based on how many participants selected them (out of a total nine) into *frequently used, somewhat used, and seldom used.* The interpreted valence of each word as well as its number of occurrences is also indicated. Words selected by less than two participants are omitted here

form of bias. However, the main takeaway from both assessments appears to be that the participants found the interface straightforward and easy-to-follow, but that its limitations in terms of tracking accuracy and expressive potential negatively contributed to their overall perceptions of its usability.

This work had several limitations, both in terms of design and evaluation. The interaction design provided only limited parametric control over the physical model properties. Future versions should enable a greater degree of user control over expressive parameters related to articulation, strike position, timbre, and dynamics, aside from adding a wider selection of instruments and possibly step sequencing functionality like in past DMIs [7, 8]. The parameter mappings should explore the use of embodied conceptual metaphors as a means to create meaningful movement-sound links [24]. In terms of technical considerations, support should be added for several eye-tracker models, and rigorous assessments of computational load and tracking accuracy should be conducted to ascertain the pros and cons of the various eye-tracker morphologies [5]. Our evaluation assessed usability and desirability with convenience-sampled participants, but did not investigate user perceptions of the expressive capabilities of the system with the user group. An evaluation procedure similar to that done in [7] is a good reference for future versions of our framework. Lastly, an interdisciplinary approach that involves end-users (e.g, ALS patients) in the design process would be the ideal approach to ensure that the developed system services its target audience.

7. CONCLUSIONS

In this work, we were able to establish a feasible digital musical interface that integrated gaze tracking with real-time physical modelling synthesis in a scalable, open-source prototype. Whilst the interface was found to be a feasible means for gaze-driven music performance our qualitative results indicate that the utility of the interface can be enhanced by expanding the possibilities for expressive control over the physical model There is indeed much room for further development, improvement, and evaluation, but we believe that the present work provides a robust technical framework for the future exploration of gaze-controlled physical models for musical expression. We hope that this can serve as a step towards enabling individuals with severe physical impairments to experience a rich world of digitally augmented creative musical experiences.

Acknowledgments

DK was responsible for system design, development and evaluation. PK assisted with technical implementation and SS supervised the project. All authors collaborated on writing the manuscript, and approved of the final version. We would like to thank the usability test participants for their time and inputs. This work is partially funded by NordForsk's Nordic University Hub, Nordic Sound and Music Computing Network NordicSMC, project number 86892.

8. REFERENCES

[1] J. Shepherd and P. Wicke, *Music and Cultural Theory*. Polity Press ; Published in the USA by Blackwell, 1997.

[2] J. Malloch, D. Birnbaum, E. Sinyor, and M. M. Wanderley, "Towards a new conceptual framework for digital musical instruments," in *Proceedings of the 9th international conference on digital audio effects*. Citeseer, 2006, pp. 49–52.

[3] F. Grond, K. Shikako-Thomas, and E. Lewis, "Adaptive musical instruments (amis): Past, present, and future research directions," *Canadian Journal of Disability Studies*, vol. 9, no. 1, pp. 122–142, 2020.

[4] E. Frid, "Accessible digital musical instruments—a review of musical interfaces in inclusive music practice," *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 57, 2019.

[5] P. Majaranta and A. Bulling, "Eye tracking and eye-based human–computer interaction," in *Advances in physiological computing*. Springer, 2014, pp. 39–65.

[6] L. C. Wijesekera and P. N. Leigh, "Amyotrophic lateral sclerosis," *Orphanet journal of rare diseases*, vol. 4, no. 1, pp. 1–22, 2009.

[7] R. Ramirez and Z. Vamvakousis, "The eyeharp: A gaze-controlled digital musical instrument," *Frontiers in Psychology*, vol. 7, 2016.

[8] W. Payne, A. Paradiso, and S. K. Kane, "Cyclops: Designing an eye-controlled instrument for accessibility and flexible use," 2020.

[9] S. Greenhill and C. Travers, "Focal: An eye-tracking musical expression controller." in *NIME*, 2016, pp. 230–235.

[10] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*. Logos Verlag Berlin, 2011.

[11] C. Abras, D. Maloney-Krichmar, J. Preece *et al.*, "User-centered design," *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, vol. 37, no. 4, pp. 445–456, 2004.

[12] N. Davanzo and F. Avanzini, "Hands-free accessible digital musical instruments: Conceptual framework, challenges, and perspectives," *IEEE Access*, vol. 8, pp. 163 975–163 995, 2020.

[13] S. Valencia, D. Lamb, S. Williams, H. S. Kulkarni, A. Paradiso, and M. Ringel Morris, "Dueto: Accessible, gaze-operated musical expression," in *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 2019, pp. 513–515.

[14] J. O. Smith III, "Viewpoints on the history of digital synthesis," in *Proceedings of the International Computer Music Conference*. International Computer Music Association, 1991, pp. 1–1.

[15] S. Bilbao, *Numerical sound synthesis: finite difference schemes and simulation in musical acoustics*. John Wiley & Sons, 2009.

[16] A. Chaigne and V. Doutaut, "Numerical simulations of xylophones. i. time-domain modeling of the vibrating bars," *The Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 539–557, 1997.

[17] J. L. Moore, "Acoustics of bar percussion instruments," Ph.D. dissertation, The Ohio State University, 1971.

[18] M. Kumar, J. Klingner, R. Puranik, T. Winograd, and A. Paepcke, "Improving the accuracy of gaze input for interaction," in *Proceedings of the 2008 symposium on Eye tracking research & applications*, 2008, pp. 65–68.

[19] M. Kumar, "Guide saccade detection and smoothing algorithm," *Technical Rep. Stanford CSTR*, vol. 3, p. 2007, 2007.

[20] J. R. Lewis, "The system usability scale: past, present, and future," *International Journal of Human–Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.

[21] J. Brooke, "Sus: a "quick and dirty'usability," *Usability evaluation in industry*, vol. 189, 1996.

[22] J. Benedek and T. Miner, "Measuring desirability: New methods for evaluating desirability in a usability lab setting," *Proceedings of Usability Professionals Association*, vol. 2003, no. 8-12, p. 57, 2002.

[23] M. Toivanen, "An advanced kalman filter for gaze tracking signal," *Biomedical Signal Processing and Control*, vol. 25, pp. 150–158, 2016.

[24] S. Roddy and B. Bridges, "Addressing the Mapping Problem in Sonic Information Design through Embodied Image Schemata, Conceptual Metaphors, and Conceptual Blending," *Journal of Sonic Studies*, no. 17, 2018.