Xyborg: A Wearable Hand-based Instrument for Musical Expression

Kristian Eicke
Department of Musicology
University of Oslo
Oslo, Norway
kristeic@uio.no

Stefano Fasciani
Department of Musicology
University of Oslo
Oslo, Norway
stefano.fasciani@imv.uio.no

Çağrı Erdem
Department of Informatics
University of Oslo
Oslo, Norway
cagrie@ifi.uio.no

ABSTRACT

This paper describes the design, implementation, and evaluation of Xyborg, a wearable instrument designed to leverage hand input gestures and arm movements. Xyborg's aims include developing a low-cost musical instrument, providing the performer with expressive control over sound parameters, and establishing a transparent connection between gestures and the resulting sounds for spectators. The design of this instrument prioritizes controllable sonic features and a fixed comprehensible mapping between gestures and sounds, rather than aiming for a broad range of functions and variable mapping. The participants in the evaluation unanimously agreed on the potential for long-term engagement and recognized Xyborg as an instrument that enables expressive sound control through movement, emphasizing its skill-based nature.

Author Keywords

hand controller, interactive music system, musical interface

CCS Concepts

•Applied computing → Sound and music computing; Performing arts; •Human-centered computing → *Interaction design;*

1. INTRODUCTION

Xyborg is a novel wearable musical instrument, designed with traditional instruments and desktop controllers in mind yet focusing on a more physically engaging and bodily playing experience by offering sound generation and manipulation using arm, hand, and finger gestures. Unlike similar instruments, which are often tailored for solo performances or specific compositions [1], Xyborg blends controllability with a familiar fingering layout and material design, aiming for intuitive use in live ensemble performances and collaborative musicking.

Xyborg focuses exclusively on real-time sound synthesis and is conceptualized to be a standalone instrument that is low-cost, open-source, and relatively easy to fabricate. Technical documentation and source-code are available online [1]. Capacitive touch and accelerometer sensors enable users to produce musical notes across a chromatic scale and control synthesis parameters. This paper outlines Xyborg's design philosophy that explores embodied, gesture-driven sound expression along with reproducibility of musical tasks, drawing inspiration from Michel Waisvisz's *The Hands*, which similarly highlighted a "balance between structure and physicality" [2].

2. RELATED WORKS

Monitoring arm, hand, and finger movement has been a prominent goal in Human-Computer Interaction since the late 1970s [3]. The latter decade saw the invention of remarkable prototypes, including usage of the then brand-new MIDI protocol [4] and real-time sound synthesis [5]. Since then, numerous systems have been designed, with some of them being transformed into musical general-purpose controllers such as Imogen Heap's *mi.mu gloves*[2][6] or the *genki ring*[3]. These instruments are gestural controllers, in essence, asking the user to implement their own mapping [7].

Other systems have remained in a state of continuous development by their respective designers or performers in favor of individualization. Laetitia Sonami's *lady's glove*[4], Michel Waisvisz's *The Hands*[5][4] or the virtuosic beatjazz explorations with self-build controllers by Onyx Ashanti [6] are such examples that allow controlling sound synthesis and sample manipulation. Similarly, there are a number of interfaces mainly designed as controllers [8][9]. However, such designs often tend to be inaccessible to other musicians and artists.

In particular, Michel Waisvisz's instrument *The Hands* is an iconic milestone in the field of wearable digital musical instruments. Waiswisz's first simple yet effective hardware controller design of *The Hands* has been determined to be a good starting point to base Xyborg's design off of in order to explore control and embodiment. Version 1 of *The Hands*, realized in 1984 and visible in Figure 1, consists of a pair of data gloves constructed with metal plates and equipped with elastic bands for secure fastening. The gloves housed multiple sensors, such as momentary push keys, mercury tilt switches for orientation data, and an ultrasonic receiver and transmitter for distance sensing. Each hand has 12

[1]https://github.com/shx-vhs/xyborg
[2]https://mimugloves.com/
[3]https://genkiinstruments.com/
[4]https://sonami.net/portfolio/items/ladys-glove/
[5]https://cracklemusic.org/TheHands.html
[6]https://www.ted.com/talks/onyx_ashanti_this_is_beatjazz

pitch keys, which are manipulated by the index, middle, ring, and little finger. Similar to *The Hands*, Xyborg also uses the full arm movement potential [10].

Concerning mapping strategies and the software side of instrument design, Xyborg relates to a more general design philosophy found in Waiswisz' work rather than specific technical aspects of *The Hands*. We refer, in particular, to the idea of balancing structure and physicality [2]. Xyborg's design aims to strike a balance between commercial gestural controllers and artistic NIME projects, emphasizing precision in sound generation while allowing for expressive play.

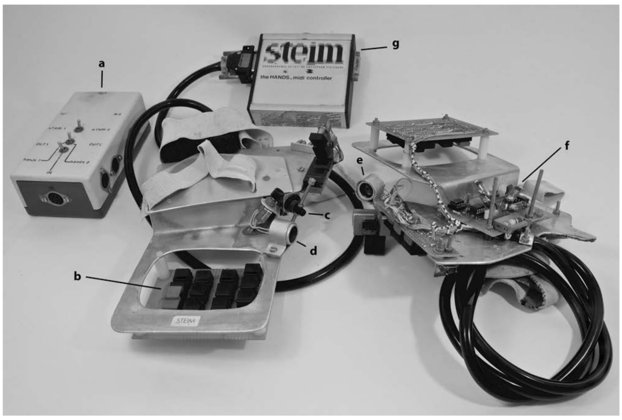

Figure 1: Michel Waiswisz' first prototype of The Hands from 1984.

3. DESIGN AND IMPLEMENTATION

3.1 Conceptual Ground

While developing Xyborg, we aimed at balancing the instrument's implications of structure and physicality by borrowing from traditional instrument metaphors while targeting the potential of full body motion "in the air". In our design, we have established three key principles:

- Access to the full range of motion of both arms.
- Expressive control of pitch and effect parameters.
- Transparent gesture-to-sound relationships.

These principles, aligned with Malloch et al.'s *skill-based interaction* framework [11], foster a transparent, intuitive connection between performer and instrument, emphasizing expression, control, and clarity. As such, Xyborg employs both limbs and complex mapping structures that allow a "high ceiling" [12] for mastery and subtle timbral control.

Given Xyborg's emphasis on performer gestures, enabling real-time expressive nuances and varied expression states is crucial, as highlighted by Dobrian [13]. That puts an emphasis on selecting appropriate sensors and crafting effective mappings. West et al. stress that a transparent mapping structure effectively conveys the performer's motion-sound congruity and illustrates the instrument's function [14]. The performer must possess the ability to deliberately influence the outcomes through their movements, which spectators should also recognize.

Incorporating metaphors or analogies can enhance control and clarity, creating a shared understanding among the instrument creators, performers, and audiences [14]. In that regard, Fels et al. argue that metaphors also improve transparency and expression [15]. Balancing mapping transparency with control complexity is vital to avoid the instrument becoming too challenging or confusing in motion–sound relationships [16]. That is particularly important to enable "flow" processes, providing instant, casual feedback while fostering a sense of discovery [17]. Such a balanced approach involves integrating embodied metaphors derived from established instrument practices and introducing complexity through a dynamic timbre space manipulation, connecting actions to timbre dimensions rather than mere synthesis parameters [14].

3.2 Physical Interface

Xyborg's specific setup allows complete mobility and independence from a laptop in a performance situation. Figure 2 illustrates an overview of the hardware layout of the instrument. In addition to that, Figure 3 and Figure 4 show the fully assembled physical interface.

We opted for custom-made capacitive pitch keys over push buttons for several reasons. Firstly, they enable the user to adjust the size and position of the gestural controller according to their preferences. Secondly, unlike push buttons, they potentially enable better agility due to the absence of any significant trigger distance. Trill Craft, a breakout board [7] for a capacitive touch sensing (CTS) integrated circuit, is used to interface the pitch keys. The sensors are placed at a relative height to provide a small degree of tactile feedback. Aluminum tape is used to wrap the bare ends of stranded wire around 20x20 mm PCB board pieces. An additional layer of adhesive tape protects the aluminum surface. Each of the twelve keys represents a step on the chromatic scale. Two pitch keys were assigned to the index and middle fingers to control six keys with four fingers. The thumb can also control functions with push buttons and CTS' in each hand. Each frame has a top-mounted triaxial analog accelerometer, which data is also used to extract the orientation of each frame. Waiswisz's initial design specified a 3x4 arrangement of the pitch keys, allowing for a complete octave to be played on each hand. However, attempting to navigate to a specific pitch key with one finger without inadvertently triggering another would be exceedingly challenging. Consequently, octave range control was outsourced to additional CTSs on the upper frame.

3.3 Sound Engine and Mapping

The sound engine of Xyborg uses wavetable synthesis to create an independent waveform for each pitch key. The signals are then combined and processed through a low-pass filter, a distortion effect based on wavefolding, and finally a modulation effect. Xyborg's implementation uses the graphical programming language *Pure Data* (Pd) running on *Bela*[8].Figure 5 illustrates the mapping and signal flow for each hand controller. The selection of effect parameters for manipulation and mapping strategies is driven by achieving a balance between complexity and transparency. Clear and unambiguous mapping is a main ingredient for establishing transparency between the performer and the spectator. However, a straightforward one-to-one mapping would not be appropriate for a skill-based interaction, and thus, it was necessary to enhance it with a more intricate approach.

To simplify the control and improve the transparency of key musical elements, we use primarily one-to-one mappings for the filter's pitch, octave control, and cutoff frequency.

[7] `https://bela.io/products/trill/`

[8] `https://bela.io/products/bela-and-bela-mini/`

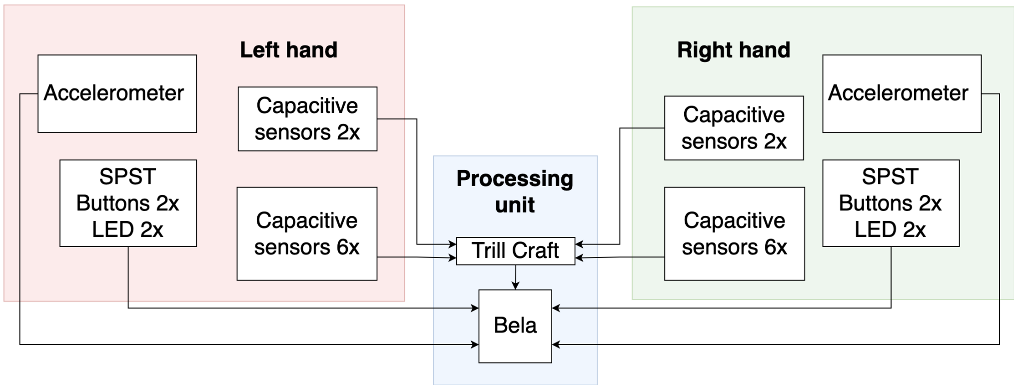

Figure 2: The hardware layout of Xyborg. The physical interface includes two wooden frames (one for each hand), two accelerometers, twelve custom-made capacitive sensors (six for each hand), and four push buttons with associated LED indicators. The *Bela* is used to implement the standalone instrument. A plastic enclosure houses the capacitive sensor board, a power bank, and the Bela board. All items are secured around the waist with an adjustable belt, making the system completely standalone and wearable. The addition of a commercial audio wireless transmitter allows for cable-free performances.

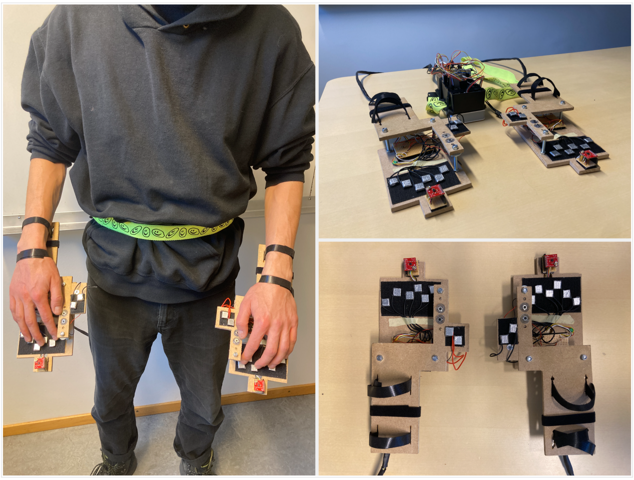

Figure 3: Xyborg worn by a performer.

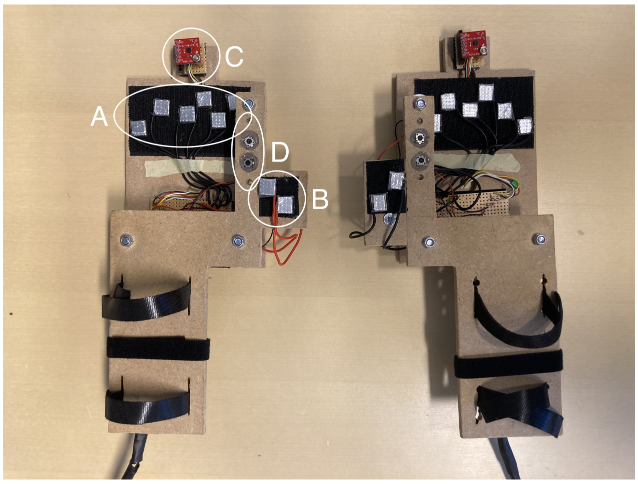

Figure 4: The fully assembled physical interface. Highlighted components include: A) pitch CTS keys, B) octave range CTS keys, C) accelerometer module, D) push buttons and LEDs

The input data from all CTSs is subjected to thresholding in Pd in order to generate a trigger. This trigger either serves as an envelope trigger or as an input for the octave range. Each individual CTS is assigned a pitch note. By adding two more touch sensors, each hand gains the ability to adjust its pitch range independently. Xyborg has a total range of four octaves, and the note C is assigned to the little finger of the left hand. The push buttons on both hands are assigned to activate or deactivate the effects. One of the buttons on the left side activates distortion, while the buttons on the right side control the filter and modulation effect. One button is currently inactive. Each LED indicates an effect's activation status.

The filter's cutoff frequency is directly mapped to the pitch of the right hand (y-axis of the the accelerometer). Raising the hand directly increases the filter cutoff, resulting in a sonic expansion. The filter is completely closed when the arm is in a resting position. Many-to-many mappings are used for the two other effects. The modulation is controlled by the cumulative acceleration data of the left hand along the three axis, which determines the modulation's depth and rate of oscillation. The speed at which the user performs a modulation motion, resembling that of a string instrument, directly impacts the intensity and tempo of the modulation. The distortion effect includes three consecutive wavefolders, with a total of 7 controllable parameters. In this case, we do not use an explicit mapping strategy but opted for a generative technique, based on an artificial neural network. This approach helps in facilitating reproducibility and a sense of embodiment during performance. Fiebrink and Caramiaux present methodologies and fundamental concepts for leveraging a machine learning algorithm as a means of developing creativity [18]. The Pd framework *neuralnet*[9] has greatly simplified, accelerated, and enhanced the procedure of creating a satisfactory distortion parameter mapping [19] using a artificial neural network as a multi-output regressor. The network architecture includes an input layer that takes the accelerometer data from the left hand, two hidden layers with 64 neurons each and sigmoid activation functions, and an output layer with linear activation functions. The output layer produces 7 values that control a global folding amplitude parameter and 6 boundaries for lower and upper clipping of the distortion effect. The training was conducted iteratively on a separate Pd patch, employing "anchor" examples [18]. These anchor examples are accelerometer axis data triplets, which repre-

[9]https://github.com/alexdrymonitis/neuralnet

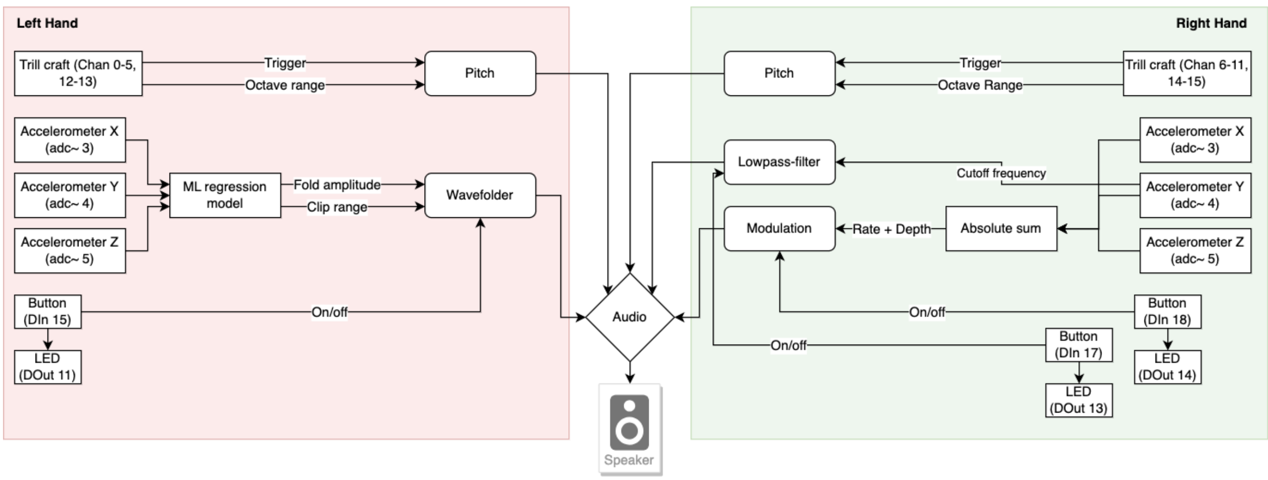

Figure 5: Diagram summarizing the mapping of left and right hand interfaces to the sound engine of Xyborg.

sent the orientation of the hand in relation to the Earth's gravitational field. This makes it possible to investigate the complete range of arm and hand movement to discover novel sounds that exist both within and outside the defined anchor examples. The pairs of axis data to distortion parameter were created in a way that a totally relaxed arm and hand position translates to *no distortion*. The effect gets more pronounced and shifts to different timbres as the hand and arm are raised and twisted.

4. EVALUATION

Wanderley and Orio provide a methodology for evaluating the usability of digital musical instruments [20]. They recommend that evaluations should be constructed around simple musical tasks. In choosing appropriate tasks for usability evaluation, they suggest that relevant features to be tested might include explorability, reproducibility, and feature and timing controllability. Two studies have been conducted: one user exploration session with a questionnaire and one with a spectator questionnaire. Both questionnaires included seven-point rating scales and questions that participants could answer freely [21].

In the preliminary user exploration, three participants evaluated the system for half an hour each. All of them felt very comfortable playing an instrument that involved pressing keys or buttons. The evaluation sessions comprised a brief introduction to the instrument's key functionality, followed by the execution of musical tasks and a survey.

These tasks were included to shed light on feature and timing controllability and reproducibility of events. The survey comprised 2 personal questions, 7 questions to rate the easiness of execution and repeatability of specific musical control tasks on a scale from 1-7 (1 hardest, 7 easiest) , and 5 open questions related to ergonomics and functionality. Finally, the participants were allowed to play around with the instrument freely and, if they wanted, perform along a rhythmic backing track. Table 1 shows each participant's musical task rating results.

A second study was conducted afterward, focusing on how spectators perceived the performance. No demographics or information on prior knowledge about NIME concepts has been collected. A total 7 participants answered the questionnaire. The survey did not focus on a specific target group and asked the participants to watch a 3-minute video of a performance with Xyborg, which is available online[10].

[10] https://mct-master.github.io/interactive-music/2023/12/04/kristeic-xyborg.html.

Attribute			
Single consecutive notes	6	6	3
Intervals	4	5	5
Arpeggios	5	6	3
Chords	5	6	6
Filter	7	7	4
Modulation/ Flanger	7	7	3
Distortion	7	6	4

Table 1: Participants responses rating easiness of execution and repeatability of specific musical control tasks.

Question							
Q1	7	6	6	7	5	7	7
Q2	6	5	6	5	4	7	6
Q3	7	5	4	7	5	7	6

Table 2: Participants rating on different performance aspects. Q1) How engaging was the performance for you? Q2) Did you find it easy to understand the relationship between the performer's gestures and the sound produced? Q3) To what extent did you feel the performer effectively conveyed expressive nuances through his gestures?

The questionnaire included 4 open questions on engagement as well as 3 rating questions on a scale 1-7 on the clarity of the gesture-to-sound relationship (7 clearest). Table 2 shows the results of the rating.

5. DISCUSSION

5.1 Hardware Layout and Controllability

The primary areas for improvement are ergonomics and the overall arrangement of the sensors. It became increasingly evident during practice and exploration that it is difficult to fasten the device onto your arms without assistance. This poses an obstacle to practicing and requires redesign.

Although the placement of the pitch keys was liked, the metaphor of an increasing chromatic scale over two hands, like on a traditional piano layout, was rated unintuitive. One participant related to that:

> "I liked the placement of the buttons for pitch control. However, the placement of the octave up/down buttons was a bit confusing because they could be confused with the pitch buttons. Having different octaves on different hands is interesting because it allows you to play two melodies at the same time. However, having half an octave for each hand is not very intuitive."

Another participant found the octave range buttons useful but explained that since there were four of them, they required substantial extra concentration to make them worthwhile musically. Moreover, the exclusion of the thumb for controlling a pitch key is a significant limitation. This, in conjunction with the key configuration for the remaining fingers, ultimately undermines the analogy of the piano keyboard layout and necessitates an exceptionally high level of concentration to execute the desired notes. In fact, this aligns with the argument that deviating from a mapping analogy can actually decrease transparency [15]. An alternative approach to design must be explored to enable the thumb to operate a pitch key while retaining access to the current functionality. Furthermore, it will be necessary to establish a pitch key layout that is evenly distributed among the remaining four fingers. The capacity to personalize the pitch key location on the board is a valuable attribute in relation to ergonomics and will be retained.

The ease of executing all musical tasks and the overall controllability of effect parameters were rated positively. Still, executing notes accurately and rhythmically remains a challenging endeavor. The design of the frame and the key layout restrict hand movement and have potential for optimization. Despite the fact that the tempo of the backing track was intentionally set to a very slow pace of 65 BPM, each user struggled to keep up with the timing while playing along. Consequently, better timing controllability has to be implemented and tested with the next iteration.

5.2 Expressive Control

Overall, the feedback of Xyborg was positive, with participants finding the instrument engaging to watch and play. In the user exploration, everyone felt they could express musical ideas with the instrument. The filter effect was found to be the most intuitive effect. The effectiveness of conveying expressive nuances through gestures was rated with a 6 on average. While practicing, the act of adjusting effect parameters and navigating through the timbre space felt instinctive and allowed for a high level of expressiveness. The participants in the study unanimously agreed and expressed appreciation for the approach of modulating the overall sound and dynamics through their movements. By employing a regression model to map the distortion parameter and adopting a straightforward mapping strategy for the other two effects, it is possible to integrate the entire range of motion for both arms effectively.

5.3 Transparency

Spectators mostly found it easy to understand the gesture-to-sound relationship. However, abrupt changes in timbre were found not to link clearly to a gesture. The relatively small gestures for changing pitch were often overshadowed by general hand movement and were, therefore, hard to link. Interestingly, one participant pointed out that pauses in sound were confusing. Half of the participants could not make a clear connection between a specific hand and the effect on this hand. Nevertheless, the mapping strategy demonstrates transparency and an adequate level of complexity. Spectators who watched the performance found it engaging and appreciated the instrument's expressive capabilities. This accounts for the stable and transparent communication of gestures aimed for as a design principle.

Addressing the inconsistent analogy for the pitch keys and their mapping can improve the mapping transparency for the performer, which would significantly reduce the level of concentration required to perform. The spectator's inability to discern which effect was manipulated by each hand is attributed to a deficiency in performance technique rather than an inherent flaw in Xyborg's design. This is in line with the statement that the disintegration of the gesture-to-sound relationship can either be caused by mistakes by the performer or instrument failure [21]. Due to limited rehearsal time, the performer was unable to fully demonstrate the instrument's capabilities, resulting in restricted and at times clumsy movements. Nevertheless, two participants expressed that this did not affect them enjoying the performance, since there remained a distinct correlation between movement and sound.

6. CONCLUSION

Xyborg was developed to investigate the expressive potential of a wearable instrument inspired by a highly recognized and influential model in the field. It demonstrates the legitimacy of investigating the connections between gestures and sound in a live performance setting without relying on prerecorded material. Two evaluation studies helped in identifying the achievements and deficiencies of the current system. Adequate controllability has been attained, yet there is scope for enhancement, particularly in regard to play speed and timing. This is partly due to the unsatisfactory ergonomics and the unnecessarily intricate pitch key analogy. Nevertheless, Xyborg allows for expressive control and demonstrates sufficient transparency in gesture-to-sound relationships for the performer and spectators. The authors see Xyborg as a valuable contribution in exploring factors of structure, such as a straightforward mapping strategy, while at the same time not compromising too much of desired physicality. Possible future enhancements may involve the redesign of the frame to address ergonomic disadvantages and the refinement of pitch key allocation. In the next iteration, the redesign involves eliminating the top part of the Xyborg and placing it at a 90-degree angle, attaching it to the side surface of the base plate. The entire support structure could be relocated from the lower section of the arm to a more advanced configuration positioned above the arm. This may involve incorporating a flexible joint that links the wrist to the palm of the hand, enabling movement of the wrist. In addition, the thumb would be able to access both pitch keys and existing on/off functionality. In conclusion, the insights gained through the design process, as well as the subsequent evaluation, support sustained commitment to and long-term engagement with Xyborg. Improvement of the current design, enhancement of functionality, and refinement of practical skills will be the main focus for future iterations.

7. ETHICAL STATEMENT

The study with human participants was conducted in accordance with the ethical and data protection guidelines set forth by the University of Oslo. The participants were informed that their responses would remain anonymous and could be used for research purposes. No video or audio was recorded during the study.

The prototype instrument was built using electronic components provided by the University of Oslo, all of which were reused from previous projects. Upon completion of this study, these electronic components and the Bela system will be repurposed for educational purposes and future research endeavors. The authors report no conflicts of interest in this project.

8. REFERENCES

[1] Cagri Erdem and Alexander Refsum Jensenius. RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 477–482, Birmingham, UK, 2020.

[2] Volker Krefeld and Michel Waisvisz. The Hand in the Web: An Interview with Michel Waisvisz. *Computer Music Journal*, 14(2):28–33, 1990.

[3] David J. Sturman and David Zeltzer. A Survey of Glove-based Input. *IEEE Computer graphics and Applications*, 14(1):30–39, 1994.

[4] Michel Waisvisz. The Hands: A Set of Remote MIDI Controllers. In *Proceedings of the International Computer Music Conference*, Burnaby, BC, Canada, 1985.

[5] Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A Hand Gesture Interface Device. *ACM SIGCHI Bulletin*, 18(4):189–192, 1987.

[6] Thomas Mitchell and Imogen Heap. SoundGrasp- A Gestural Interface for the Performance of Live Music. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, 2011.

[7] M.M. Wanderley and P. Depalle. Gestural Control of Sound Synthesis. *Proc. IEEE*, 92(4):632–644, April 2004.

[8] Elena Jessop. The Vocal Augmentation and Manipulation Prosthesis (VAMP): A Conducting-Based Gestural Controller for Vocal Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Pittsburg, PA, USA, 2009.

[9] Yongki Park, Hoon Heo, and Kyogu Lee. Voicon: An Interactive Gestural Microphone For Vocal Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Ann Arbour, MI, 2012.

[10] Giuseppe Torre, Kristina Andersen, and Frank Baldé. The Hands: The Making of a Digital Musical Instrument. *Computer Music Journal*, 40(2):22–34, 2016.

[11] Joseph Malloch, Jérémie Garcia, Marcelo M. Wanderley, Wendy E. Mackay, Michel Beaudouin-Lafon, and Stéphane Huot. A Design Workbench for Interactive Music Systems. In Simon Holland, Tom Mudd, Katie Wilkie-McKenna, Andrew McPherson, and Marcelo M. Wanderley, editors, *New Directions in Music and Human-Computer Interaction*, pages 23–40. Springer International Publishing, Cham, 2019.

[12] David Wessel and Matthew Wright. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal*, 26(3):11–22, 2002.

[13] Christopher Dobrian and Daniel Koppelman. The 'E' in NIME: Musical Expression with New Computer Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, 2006.

[14] Travis West, Baptiste Caramiaux, Stéphane Huot, and Marcelo M. Wanderley. Making Mappings: Design Criteria for Live Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Shanghai, China, 2021.

[15] Sidney Fels, Ashley Gadd, and Axel Mulder. Mapping Transparency Through Metaphor: Towards More Expressive Musical Instruments. *Org. Sound*, 7(2):109–126, 2002.

[16] Andy Hunt, Marcelo M Wanderley, and Ross Kirk. Towards a Model for Instrumental Mapping in Expert Musical Interaction. *ICMC*, 2000.

[17] Mihaly Csikszentmihalyi. The Flow of Creativity. In *Creativity: Flow and The Psychology of Discovery and Invention*, volume 39, pages 95–112. Harper Collins, New York, NY, USA, 1996.

[18] Rebecca A. Fiebrink and Baptiste Caramiaux. The Machine Learning Algorithm as Creative Musical Tool. In Roger T. Dean and Alex McLean, editors, *Oxford Handbook on Algorithmic Music*, volume 1. Oxford University Press, 2018.

[19] Alexandros Drymonitis. [neuralnet]: A Pure Data External for the Creation of Neural Networks Written in Pure C. In *Proceedings of the International Conference on AI and Musical Creativity*, Brighton, UK, 2023.

[20] Marcelo Mortensen Wanderley and Nicola Orio. Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal*, 26(3):62–76, 2002.

[21] Sile O'Modhrain. A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal*, 35(1):28–42, 2011.