SynthSign – An American Sign Language Controlled Synthesizer with a Laser Projection Display

Tristan Barber, Jacob Goc, Christian Paredes, Christopher Jean

University of Central Florida, College of Electrical and Computer Engineering, College of Optics and Photonics, Orlando, FL

***Abstract* — This paper documents the design and testing of SynthSign, including software, hardware, and optical elements. SynthSign is an electronic musical device that allows users to synthesize sounds using American Sign Language (ASL) gestures in front of a camera. Additionally, the device contains a Laser Projection Unit (LPU) which acts as the graphic user interface (GUI) for the user.**

***Index Terms* — Photonics, Laser, Lens, Cameras**

I. INTRODUCTION

The purpose of this project is to design an all-in-one solution for music hobbyists to create and enjoy music, without the issues brought about by traditional equipment. The project will be highly technical in nature, with design requirements in software, hardware, and optics. Primarily, the device will use computer vision to detect gestures from a user and play various tones based on the gesture provided. The project will also have a graphic user interface via a laser projector, which will display information to improve the user experience.

A. Goals and Objectives

Our core goals include the ability to synthesize any of the 12 notes within an octave. This collection of notes is called a chromatic scale. We also want to provide a three-octave range, from notes C3 to B5. As this synthesizer will be controlled using hand gestures, a pivotal goal is to develop a machine learning object detection model to process user gestures. We also want the device to be able to interface with Bluetooth speakers.

The advanced goals we wish to achieve exist to provide the team with some exciting challenges. We aim to extend our octave range from three to five, giving the ability to synthesize notes from C2 to B6. Finally, we want to include a secondary mode, where users can select any one of the twelve major keys, and play preprogrammed chords based on the key signature selected.

Finally, our stretch goals. The team hopes to design a singular PCB to house all our electronics, doing away with any kind of off-the-shelf development kit. Additionally, we would like to give users the option to select between the twelve minor keys, as opposed to limiting them to major keys. We would also like to add a sustain functionality, which would imitate the sustain pedal on a piano. Allowing the user to map their own gestures would also be a fantastic feature. Finally, to improve musical performance capabilities, the team wants to develop a mechanical swivel interface that tracks the user as they move around the room.

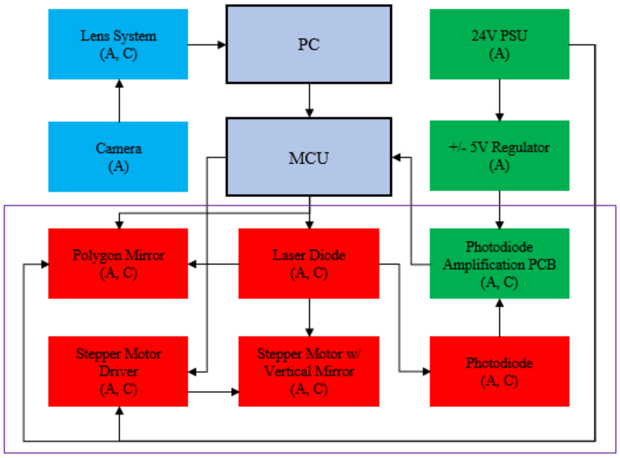
To achieve our core goals, the team has come up with a few key objectives. First, we have determined that our machine learning object detection model will need to recognize 13 distinct gestures. One gesture for each of the 12 notes in a chromatic scale, and a gesture to change the octave range Additionally, we will have to implement a sound generation library within the software.

For our advanced goals, we'll need to broaden our scope. We will need to implement an additional 8 gestures. We need a gesture to select the key signature, and a gesture for each of the seven chords within a key signature. We will also need to expand our sound library to allow for the increased octave range.

Finally, we have a few key objectives for our stretch goals. We'll need to implement facial tracking within our machine learning model, which we will use to track a user's movement. Additionally, we will have to research servo and stepper motors to build our rotating platform. Finally, we will have to add additional gestures to change the laser projection unit display color and to toggle the tracking mode.

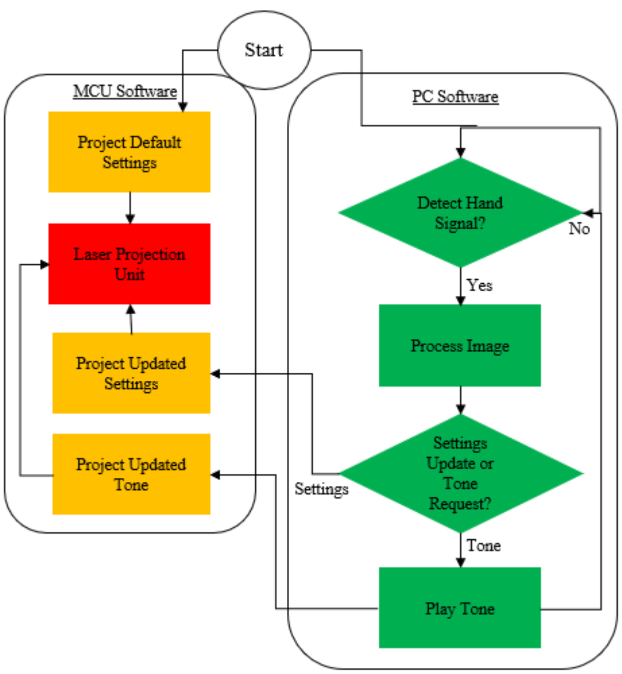
B. System Flowcharts

The following figure shows a block diagram created to provide a visual representation of the hardware components and how they work amongst each other. Additionally, each block provides detail about who is responsible for each component, and what phase of development the component is in. The two centralized blocks, PC and MCU, exist to show how and where data from individual hardware is routed to be processed by software. Arrowheads are used to show data flow between different areas of hardware. Like the software breakdown, members of the group assigned themselves tasks based on areas of personal interest and confidence.

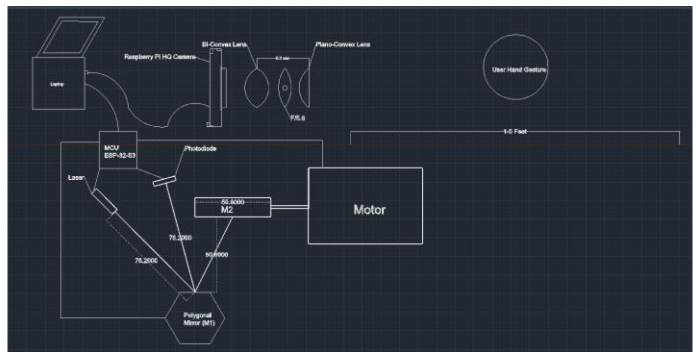


*Purple box indicates the overall Laser Projection Unit (LPU)

The following figure shows the software block diagram. The goal of this figure is to detail each component within the software, and how each component interfaces with one another to form the complete system. As is evident from the figure, the software breaks up into two key development areas. The image processing computation takes place on the PC and includes gesture recognition and sound generation. The other portion of the software, being implemented onto the microcontroller development board, controls the laser projection display, which acts as the graphic user interface for the device. Additionally, it will control the variable zoom lens based on data.



For the optical diagram, we will discuss the optical schematic diagram that will be further dissected in the following sections later in the report. The project itself will consist of two optical designs provided by our following photonics members. It will comprise the two-lens system and the laser projection unit's design. The lens system will comprise of the raspberry pi high quality camera, followed by a bi-convex and plano-convex lens. It contains a physically structured aperture to help further enhance our distance requirement. The laser projection unit itself will involve the utilization of an instapark DRM104-D003 laser diode, with a photodiode and a controlled mechanism. This mechanism will consist of an aperture and a polygonal mirror that is driven by a DC motor. Finally, the projection unit will output the resulting display. Below, the schematic diagram illustrates the two optical designs highlighted to help showcase where each lie in respect to the overall diagram.



II. ENGINEERING SPECIFICATIONS

Below is a table of our three key engineering specifications.

Requirement	Specification
Gesture Detection Accuracy	Shall meet or exceed 80% of captured gestures.
Response Time	Less than or equal to 2 seconds
Recognition Distance	Less than or equal to 5 feet

III. LASER SAFETY STANDARDS

A large part of this project involves the use of a low power laser around the 633nm wavelength. Although this is not a particularly dangerous laser, there are still risks associated with it. One thing we do not have to worry about is the skin damage, since our laser is in the visible range and is only a few mW in power. However, any amount of eye exposure is dangerous and can cause permanent damage, no matter the power of the laser. Even if the lasers expressed intent is to be pointed towards a surface to create a GUI, accidents can still occur, so knowing the safety standards behind lasers is important in designing the LPU.

The Laser Institute (LIA), in collaboration with the Occupational and Safety Health Administration (OSHA), created the ANSI Z136 series of laser safety standards. There are multiple subsections of the series that are pertinent to our project. Z136.1 – Safe Use of Lasers, being the parent document, lays the foundation for the standards in industry and academia. Since we are doing this project in an educational setting, and utilizing senior design labs found on campus, Z136.5 – Safe Use of Lasers in Educational Institutions is pertinent when following building codes, especially those laid out for the CREOL building. This goes hand-in-hand with Z136.8 – Safe Use of Lasers in Research, Development, or Testing, as the main purpose of the design labs is to test and develop the means for which the laser will be used.

With these standards in mind, testing can go forward on the LPU, and the risk of damage being caused in the future can be mitigated. Properly protecting against uncontrolled laser guidance with beam blocks and PPE in the form of laser safety glasses are some of the methods that will contribute to this reduction.

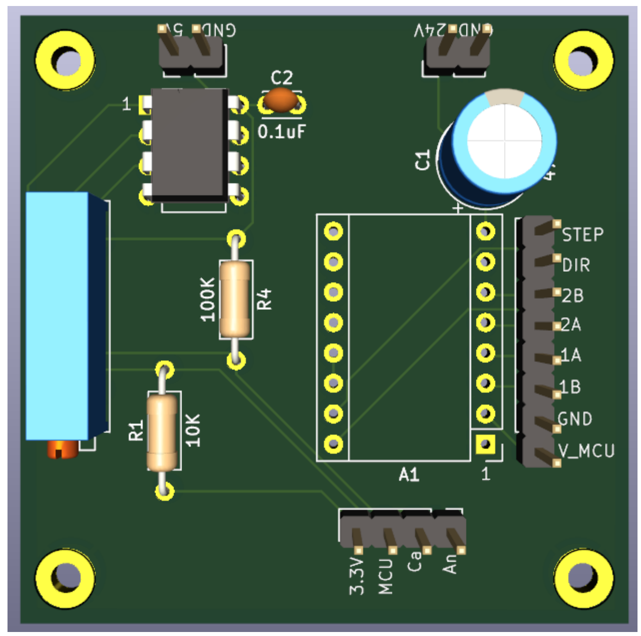
IV. HARDWARE DESIGN

Hardware Design for SynthSign is a split effort between electrical design and optical design. Generally speaking, the group was instructed to make the electrical design effort as light as possible, as there is no dedicated electrical engineering major within the team composition. Regardless, the LPU still required some electrical design, including a printed circuit board (PCB) and other miscellaneous items.

A. Electrical Design

As discussed previously, the LPU utilizes a spinning horizontal mirror to rasterize a laser across the X plane of our output. Additionally, that laser is fired into a mirror spinning on the vertical plane, to rasterize the laser in the Y direction. In order to track the beginning of a new line, a photodiode was added, which sends a signal whenever the laser crosses it. This is where some circuitry was required.

Below is our PCB. The photodiode comparator consists of an LM393 voltage comparator used to send digital data to the MCU when the laser hits the photodiode. A trim potentiometer is used to set the reference voltage on the positive input of the LM393. The negative voltage is determined using a voltage divider, with a 100K resistor and the photodiode in a reversed biased configuration. Additionally, our PCB also contains a footprint for our stepper motor controller, which was added in an effort to clean up wiring within the device. Below is a 3D model of our PCB Design:



B. Optical Design

In terms of the optical design within the hardware design, it deals with the previously mentioned LPU and with the two-lens system in conjunction with our selected camera. With our selected camera of the raspberry pi high-quality camera, we were able to gather our information regarding the sensor and its format to understand the necessary focal length needed for each field of view direction. Below is our table detailing the focal length requirement on each angular field of view.

Focal Length Requirement on Angular H/V FOV			
Angular HFOV (Degrees)	Focal Length (mm)	Angular VFOV (Degrees)	Focal Length (mm)
80	**3.746**	40	6.473
85	**3.431**	**45**	**5.688**
90	**3.144**	**50**	**5.052**
95	**2.88**	**55**	**4.526**
100	**2.638**	**60**	**4.081**
105	**2.412**	65	3.698
110	**2.638**	70	3.365
115	**2.038**	75	3.07
120	**1.815**	80	2.808

As we can see from the table, since our specification regarding field of view in the horizontal direction is 80-120 degrees, while the vertical direction is 45-60 degrees. This means ideally, we want a focal length of at least 3.746mm to ensure that the field of view is met in both directions. Since our lens design requires the utilization of a two-lens system, we utilized the formula of the compounded lens equation to garner what focal lengths would be necessary on each lens ideally. In our two-lens system, we will be utilizing a bi-convex lens, followed by a plano-convex lens. In between this configuration is the physical iris aperture that will help to further enhance our depth of field to encompass the distance of 5-feet.

Learning how to manipulate a laser beam to be moved in two dimensions was important for the LPU. Two mirrors are needed: one in the horizontal direction, and the other in the vertical direction. The first mirror will rotate upon one axis, while the second mirror will rotate on one of the other axes perpendicular to the first. For our purposes, the LPU will be designed with the polygonal mirror rotating upon the axis that is perpendicular to plane generated from the surface the GUI will be displayed, i.e., a desk, table, or any other flat surface. The second mirror will rotate upon the axis that is perpendicular to the first mirror, and the axis that passes through the user.

The laser, photodiode, polygonal mirror, and mirror-mounted motor will all be located on the same plane to reduce the amount of angular noise within the system. The second mirror will need to point downwards to enable the laser light to contact the surface upon which the LPU is sitting on.

The only thing that the distance between the laser and the polygonal mirror affects is the divergence of the laser. Minimizing this distance as much as possible reduces the amount of spread the laser light will experience, which affects how clear the GUI will be. This also affects the amount of light the photodiode receives from the laser, as the photodiode has a small active region. Any light outside of the active region will not be detected, and the intensity experienced by the photodiode will be reduced. This affects the amount of current generated by the photodiode that synchronizes the mirrors to effectively generate the text, and therefore could cause disturbances in the GUI generation.

Our specification calls for the GUI to be created at a distance of 12” from the device, with dimensions 12” long and 9” wide. The polygonal mirror will control the length of the GUI, while the second mirror controls the width. To find the angle the of the transmitted laser beam with respect to the normal of the mirror, I used $\arctan\left(\frac{x}{y}\right)$, where x is the desired width of the GUI and y is the total optical path length the laser travels. This means that, with $x = 9''$ and $y = 12''$, the laser is at an angle of 36.87°. To find the how much of an angle the motor must sweep, we divide this angle by 2, to get an angular travel of 18.44°.

Using the same equation when we calculated the angle of the second mirror, we can find that, to reach the full length of 12” at an OPL of 14”, we would need the laser to be incident on the surface of the polygonal mirror at 49.4° relative to the mirror. One thing to consider is that while the polygonal mirror is rotating, the distance between the first and second mirror will grow smaller as it approaches the edge of the hexagon the mirror is shaped out of. This should have minor effects on the size of the GUI and would be hard to correct for. As such, we will think of the specification as an average size of the GUI and not an absolute.

Since the above equation uses the OPL, the distance between the two mirrors must be considered when calculating the length provided from the first mirror. Assuming this distance is 2”, we can find not only the angle at which the laser will need to hit the first mirror, but also the determine how large the second mirror needs to be to capture the full sweep of the laser beam. This will help optimize the amount of room we have, as reducing the size of the second mirror to a little over the length we need can help reduce the width of the device.

V. SOFTWARE DESIGN

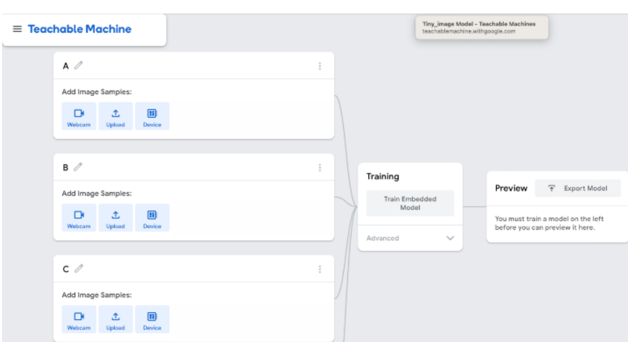
The software design effort was broken into two categories. The application software was primarily written in Python and is responsible for the gesture detection and audio output. Additionally, the application software communicates with the embedded code using UART. The embedded code is written in C/C++. It is flashed onto an ESP32 development board, which controls the pulsing of the laser diode, as well as the spinning of the motors. Essentially, the embedded code acts as the control system for the LPU.

A. Application Software Development

To follow good Python development practices, a Python virtual environment was set up on any device that was used to import a Python library. This includes the PC. A Python virtual environment is a self-contained directory that contains its own Python interpreter and a set of libraries. The purpose of using a virtual environment is to create an isolated environment for a Python project, separate from the system-wide Python installation. This helps manage dependencies, avoid conflicts between different projects, and ensure that a project can run in a consistent and reproducible environment.

To generate sound using Python, we'll have to create a couple of different Python scripts. Initially, we'll need to define a function that outputs a sine wave based on a given amplitude, frequency, and time. We'll use the pygame library for sound generation. Pygame is a powerful Python library used for making games in Python.

Our team will be using Google's Teachable Machine website to build and train our machine learning model. Teachable Machine is a powerful and intuitive tool that makes it easy to create and export a model.



A table of the hand gestures selected for use within SynthSign is shown below. Note that for spatial concerns, an example image for each gesture is not shown. That information can be found in chapter 8 of our 120-page report, for those interested.

Desired Function	ASL Gesture
Note A	Letter A
Note B	Letter B
Note C	Letter C
Note D	Letter D
Note E	Letter E
Note F	Letter F
Note G	Letter G
Accidental	Letter K
Octave	Letter O
Key Signature	Letter S
Major I Chord	Number 1
Minor ii Chord	Number 2
Minor iii Chord	Number 3
Major IV Chord	Number 4
Major V Chord	Number 5
Minor vi Chord	Number 6
Diminished vii Chord	Number 7

B. Embedded Software Development

Our gesture-controlled synthesizer project involves embedded programming that plays a pivotal role in controlling horizontal and vertical mirrors. We use C++ to manage the intricate mechanics of the laser projection unit (LPU). The microcontroller unit (MCU) is also programmed in C++ and orchestrates the movement of these mirrors, ensuring precise and synchronized motion essential for accurate graphical user interface (GUI) projection.

For the horizontal mirror, our C++ code controls a stepper motor that adjusts the mirror's angle, allowing the laser to sweep across a horizontal plane. This movement is crucial for displaying the GUI elements across the width of the projection surface. We utilize Pulse Width Modulation (PWM) signals generated by the MCU to dictate the motor's speed and direction, enabling fine control over the mirror's positioning. The embedded software calculates the required angles and velocities, translating these into specific PWM duty cycles to achieve the desired horizontal sweep of the laser beam. Similarly, the vertical mirror is controlled by another stepper motor, which is also driven by PWM signals from the MCU. The vertical mirror's movement is essential for scanning the GUI elements vertically, complementing the horizontal sweep to create a two-dimensional projection field. The C++ programming for this mirror involves precise calculations to determine the necessary motor speed and position adjustments, ensuring that the GUI is projected accurately and consistently at different heights.

Synchronizing the laser's operation and the mirrors' movement is critical in both cases. The embedded software ensures that as the laser beam moves, it precisely aligns with the GUI elements at the correct spatial coordinates. This requires a dynamic and responsive control system capable of adjusting the mirrors' positions in real time based on the laser's current location and the desired GUI layout. To achieve this high level of control and synchronization, our embedded programming includes real-time monitoring of the mirrors' positions, often employing feedback mechanisms like optical encoders or position sensors. These mechanisms provide the MCU with continuous data on the mirror's angular position, allowing for real-time adjustments and corrections to maintain the projection's accuracy.

The embedded software also incorporates safety features to prevent the mirrors from moving beyond their mechanical limits, protecting the system from potential damage. These features include limit switches or software boundaries that define the maximum range of motion for each mirror, ensuring that the stepper motors operate within safe parameters. In summary, controlling the horizontal and vertical mirrors in our synthesizer's laser projection unit relies heavily on sophisticated embedded programming in C++. This programming dictates the precise movement of the mirrors through PWM-controlled stepper motors, enabling accurate and dynamic projection of the GUI. The embedded software ensures that the mirrors contribute effectively to creating an interactive and visually engaging musical experience through meticulous calculations, real-time feedback integration, and safety considerations.

VI. TESTING AND RESULTS

Testing was also a split effort, with the same general subcategories. The three highlighted engineering specifications fell under the design of the ASL synthesizer itself, so testing for the application development focused on those three specifications. The other area of testing involved the LPU, which includes things like breadboard development.

A. Application Testing

Below is an image of testing the machine learning model for the synthesizer mode. As shown, the user is gesturing an ASL A, and the appropriate classification has been identified and placed above the hand on the image.

Much of our application testing took place during preparation for the midterm demo video. To test our gesture detection accuracy specification, we flashed each sound generating gesture onto the screen 10 times, and recorded whether or not the proper tone was output. Below is a table summarizing the 10 runs. Our average detection accuracy was 95.71%, which meets our specification of 80% or better.

Gesture:	1	2	3	4	5	6	7	8	9	10	% Correct
A:	A	A	A	A	A	A	A	A	A	A	100%
B:	B	B	B	B	B	B	B	B	B	B	100%
C:	C	C	C	C	O	C	C	C	E	C	80%
D:	D	D	D	D	D	D	D	D	D	D	100%
E:	E	E	E	E	A	E	E	E	E	E	90%
F:	F	F	F	F	F	F	F	F	F	F	100%
G:	G	G	G	G	G	G	G	G	G	G	100%

For the gesture recognition speed specification, the recording was broken down into frames, and the number of frames between the gesture being flashed and the audio being output was converted into seconds. Our specification was that audio needed to be output within two seconds. After testing, it was determined that the average gesture detection speed was approximately 0.24 seconds, so that specification was met as well.

Our third and final specification was gesture recognition distance. This specification is that gestures should be recognized within five feet directly in front of the camera. During our midterm demo, a tape measurer was set up perpendicular to the camera lens, and the gesture was flashed, then slowly moved back until recognition became inaccurate. During the midterm demo we could only achieve a maximum recognition distance of 3 feet. However, this was later rectified by updating the lens system and implementing a strategic software change, putting us in-line with that specification.

B. LPU Testing

We conducted rigorous testing and found that minimizing noise is critical for accurate signal reception within our system's constrained timing window. This was especially evident as we optimized the photodiode signal, motor drive frequency, and bit clock Hz in our Laser Projection Unit (LPU). In real-time, it was necessary to achieve precise signal capture and interpretation, underscoring the importance of noise reduction. Due to the rapid laser movements and the photodiode's swift detection of light changes, even slight noise could significantly distort signal accuracy. In the context of the LPU, where we must detect and process light interruptions quickly, maintaining a noise-free signal environment was paramount.

Multiple different circuits were tested for the LPU to run properly. Our first iteration consisted of just the photodiode and a resistor, which proved to be inadequate in the amount of voltage it provided. Our MCU was incapable of reading the small output our photodiode was providing. We then tried a circuit which utilized the reverse bias of the photodiode, which provided similar results to the previous circuit. The next iteration used two op amps in the inverting amplification configuration, which was our first major breakthrough; we were able to get readings from the photodiode large enough to generate edge detection from our MCU. However, the response time of the circuit was too slow for the speeds that our polygonal mirror would be spinning at, so another solution was needed. This leads us to our final solution.

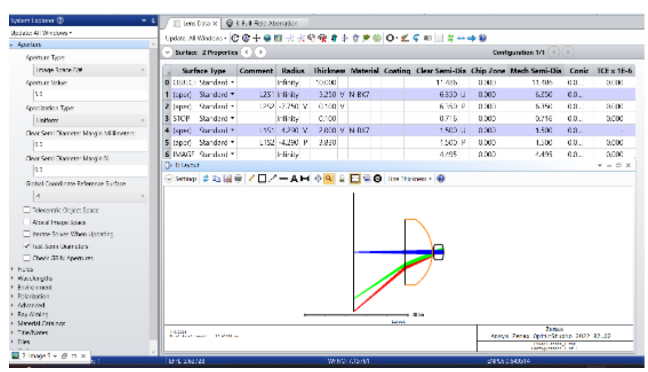
Implementing a voltage comparator was crucial in this process. It allowed our microcontroller unit (MCU) to accurately identify the photodiode's state amidst the narrow timing window. This method ensured that the system responded only to legitimate photodiode activations caused by the laser's path, eliminating false triggers from noise. Additionally, adjusting the motor drive frequency and bit clock Hz was vital for setting the system's pace, ensuring the laser's movements and signal processing were in perfect sync. This adjustment was crucial for the LPU's ability to project images and text accurately, with each pixel precisely aligned within the stringent timing requirements.

Our comprehensive testing and optimization revealed that noise reduction is essential for signal clarity and the system's overall functionality within specific timing constraints. This experience highlighted the need for a robust, noise-resistant system capable of executing complex operations accurately and reliably within a tight timeframe.

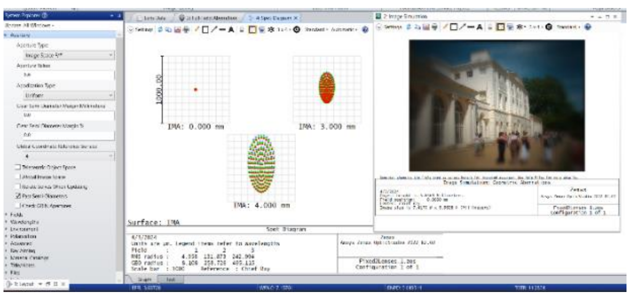
When building the prototype, and later, the final project, care was taken so that everything was lined up precisely. Multiple times the 3D printed parts needed to be adjusted and reprinted to be able to align the laser with the first and second mirror. The parts also needed to have openings large enough to fit components such as the mirror, laser diode, and photodiode, while also making sure they were not loose enough to fall out or become tilted.

C. Lens System Testing

To accurately test the potential output image regarding the entire lens system, we implemented the software simulation of Zemax. Within this application, we were able to accurately layout the order of our lenses to reach the total effective focal length necessary for reaching our angular horizontal and vertical field of view. As far as when it came to testing of the depth of field, we varied the aperture size. In the image below, we show our optical layout regarding our forementioned lens system.



In the optical layout, we flipped the way we place the plano-convex lens to have its curved side facing inwards to help with minimizing any potential aberrations from occurring. As well, our implementation of the iris aperture in between the two lenses to further help control the incoming light rays towards our sensor. We understand that due to the utilization of our two-lens system, we have had to balance the image quality and our overall field of view. Since having the plano-convex lens with its curved side facing inwards, we were able to minimize the aberrations occurring, but could not minimize the vignetting occurring on the edges of the image. However, this would not be of drastic concern since we are primarily focused on the hand gestures being captured in the center of the image regardless, as seen in the following image simulation.



VII. CONCLUSION

Despite significant challenges, our team firmly believes that we have successfully met the requirements laid out for Senior Design I and II and successfully implemented our project. Additionally, the team has paid careful attention to design scope and individual workloads, to ensure that no single member will be overloaded throughout the course of Senior Design II. We have met all of the various deadlines for senior design throughout both Senior Design I and Senior Design II.

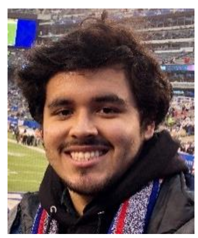
We had some struggles with part design and did suffer through an extensive redesign process during Senior Design II. Despite this, the team persevered, and did what needed to be done to get the project off the ground.

All in all, the team is immensely proud of the work we put in throughout the Senior Design pipeline and are thrilled to use the skills we've learned as we enter industry. We feel that this product could be scaled or modified to fit into many different markets, and that is an exciting idea on its own. Our ultimate goal throughout this process was to create a synthesizer that overcomes the issues of its predecessor, and we believe we made strides toward that end. In conclusion, our device combines various different features in software, hardware, and optics in order to create a successful user experience and optimize the current state of synthesizers on the market.

BIOGRAPHIES

Tristan Barber, a 20-year-old graduating with his bachelor's in computer engineering. He has accepted a position with Ford Motor Company in Sunrise, Florida, as a Product Development and Test Engineer.

Jacob Goc, a 25-year-old graduating with his bachelor's in photonic science and engineering. He has accepted a position with Everix Optical Filters in Azalea Park, Florida, as an Optical Sales and Applications Engineer.



Christian Paredes, a 23-year-old graduating with his bachelor's in photonic science and engineering. He is currently considering his future when it comes to job opportunities in the field of optics and photonics.

Christopher Jean, a 25-year-old graduating with his bachelor's in computer engineering. At present, he is considering job offers from prestigious organizations such as the Air Force and BNY Mellon, as well as exploring opportunities in start-ups.

REFERENCES

ANSI Z136 standards. The Laser Institute. (2022, April 13). https://www.lia.org/resources/laser-safety-information/laser-safety-standards/ansi-z136-standards

Iliasam. (2019, February 4). Making a DIY text laser projector. Habr. https://habr.com/en/articles/438618/