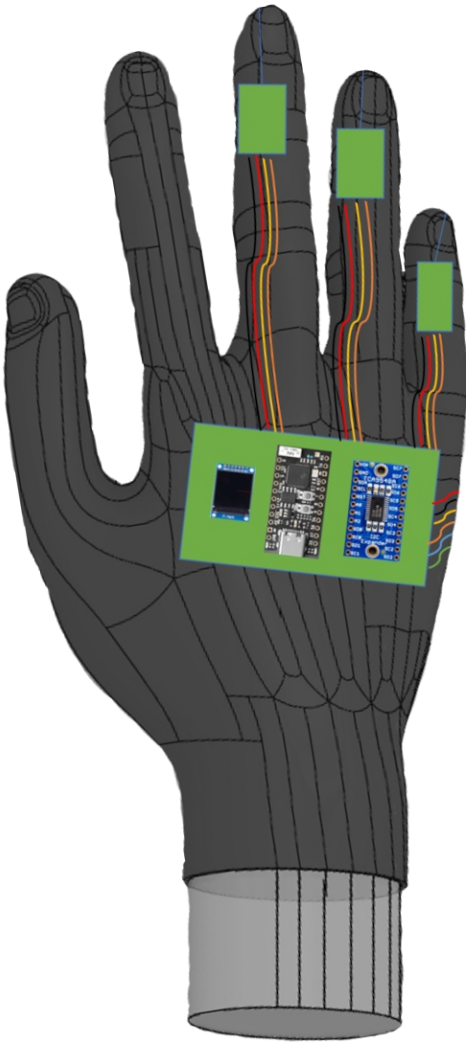
Memorandum

TO: Kimberly Lemieux (lemieuxk@camosun.bc.ca)
CC: Justin Curran (jcurran@camosun.ca);
Mel Dundas (dundas@camosun.bc.ca);
Wayne Mayes (Mayes@camosun.bc.ca)
FROM: Jose Figueroa (jose.figfatez@gmail.com);
Temo Guerrero (ccguerrero@gmail.com);
Italo Hernandez (italo.hernandez@ieee.org);
Joel Legassie (joel@jamesbay.org);
DATE: 22 Sep. 2023
RE: The Conductor: An open source device to translate human gestures into music.

Summary

We propose to develop The Conductor, a prototype glove with an embedded musical instrument controller. A microcontroller in the glove will read data from acceleration and distance sensors and transmit the data wirelessly to a PC where a neural network will recognize specific gestures and translate them into MIDI commands for a digital audio workstation (DAW). Inside the DAW, a user interface plugin (VST) will allow users to link gestures to DAW parameters and control software instruments and audio effects in real time for performance or recording.



Introduction

At the dawn of time, before Facebook and before books, even before language, already there was music. The power of music to inspire and stimulate us individually and collectively has been demonstrated innumerable times throughout the ages. We have always used the materials and technologies around us to create and share beautiful and evocative sounds.

In an age when digital electronics have put a personalised orchestra in everyone's pocket, electronic music, where the 'instrument' is code running through a processor, has become a powerful part of our repertoire. It gives artists endless options for sculpting their music, but in live performance it doesn't have exactly the same warmth and intimacy of traditional music. But now barely half a century since Wendy Carlos' *Switched on Bach* [1] made electronic music into a thing, it already, apparently, faces a challenge in the promised arrival of artificial electronic intelligences creative enough to render human musicians obsolete [2]. Could it be that electronic music is a dead end in a long tradition of human musical creativity and cooperation?

Not if The Conductor has anything to say about it.

We intend to use 'artificial intelligence' (a.i) to bring people into electronic music, not push them out. The Conductor is an embodied device that gives electronic music artists a way to physically interact with their music while they make it. Motion sensors will measure the orientation of the artist's hand and individual fingers, as well as the height of the hand above a surface. A simple a.i algorithm will recognize gestures in this data and let the user assign them to effects and filters in a Digital Audio Workstation (DAW). You can make music in real time simply by moving your hand!

The Conductor team believes that the 'electronics' in electronic music are just as much fun as the music. We aim to market the Conductor as a DIY Kit including parts, software, instructions and educational materials about electronics and electronic music. We are seeking a small initial investment of $250 to combine with a considerable amount of sweat equity to develop a working prototype, a rigorous battery of tests, and accompanying documentation.

Discussion

The Conductor will provide a new way to use existing musical tools. Generally electronic musicians use DAW software to interface with digital instruments modelled after the instruments they are already familiar with, such as keyboards and drum pads. Moog synthesisers, for example, became popular in the early days of electronic music because they made it easy for artists to bring traditional musical techniques and sounds into the electronic world [3].

The Conductor builds upon this foundation by letting electronic music artists, literally, grasp their music in a new way. They can overcome creative blocks and develop new musical ideas, while putting on a killer show.

Technical Details

From raw sensor data to musical notes streaming from a DAW, The Conductor will draw upon a chain of interconnected technologies. We have carefully chosen each

link in the chain with the goal of making live music in mind. Devices and software must be fast, accurate and tough enough to handle the rigours of live performance.

MXC4005XC Accelerometer [4]

Accelerometers are sensors that detect changes in kinetic and potential energy corresponding to changes in the sensor's velocity. The MXC4005XC is a very small device (1.18mm x 1.7mm x 0.85mm) that measures velocity changes in 3 directions and stores them in digital memory for a microcontroller to access. While it is possible to detect and collect information about almost any movement of the sensor through space, processing this data is computationally expensive. It requires precise attention to timing and careful filtering of noise. However, the data collected from the sensor can also be used to map the orientation of the sensor in relation to gravity at given moment. The values in the sensor's memory for each direction rise and fall as the angle of the force of gravity changes with the user's hand movements (see figure 1).

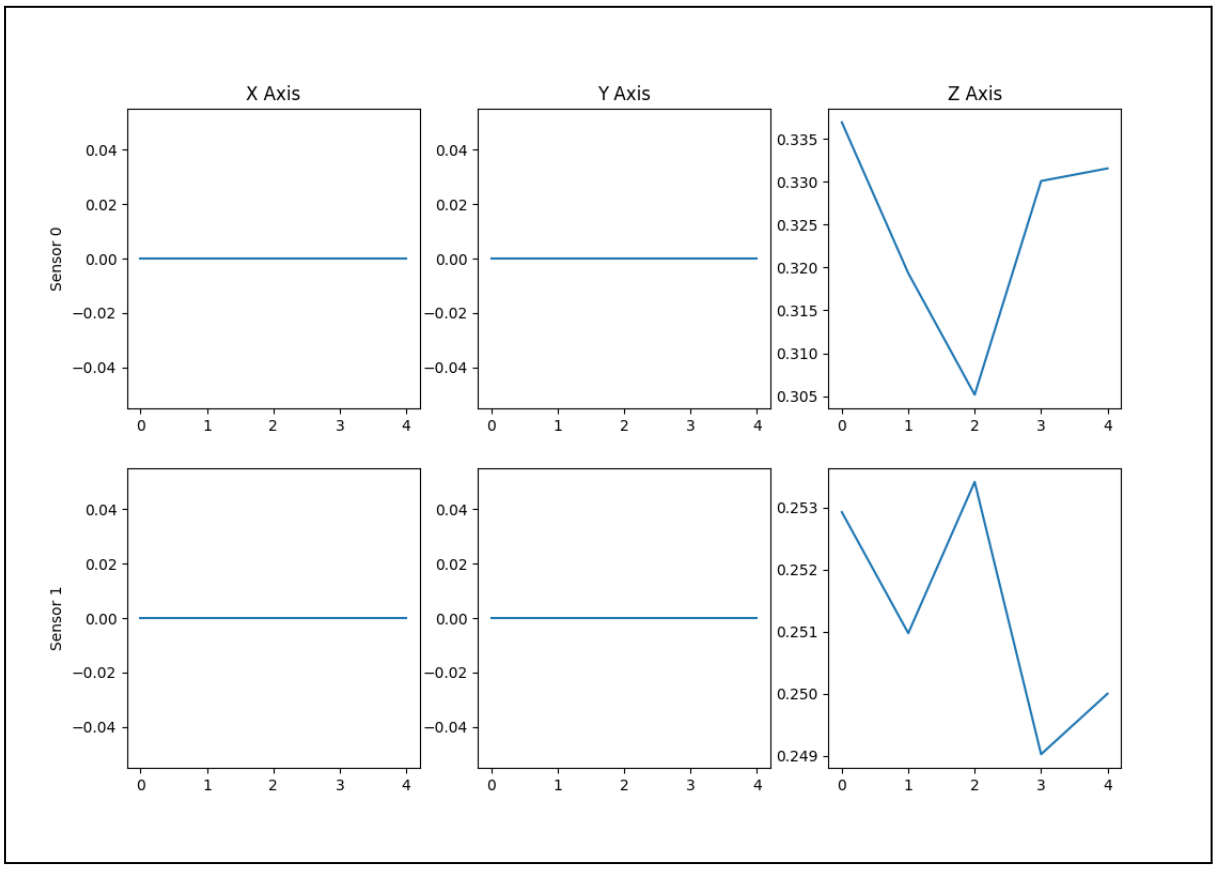


Figure 1. Visualisation of sensor data ready for input to the neural network. Each row shows the force of gravity (vertical axes) in each direction (X, Y, Z) relative to the position of one sensor over five samples (horizontal axes). The graphs show that both sensors are laying immobile with the Z direction at 90◦ to gravity. The data is almost simple enough for humans to recognize gestures in it, but the neural network does it much faster.

The Conductor will have four accelerometers, one on the back of the hand and one on each of the first three fingers. A microcontroller can harvest their data and feed it to a small and fast neural network to determine the current orientation of the sensors

relative to gravity and to each other. Sequences of sensor orientation combinations can be strung together to form gestures that will trigger effects within the DAW. The four independently moving sensors will afford a large number of potential orientation combinations from which to recognize distinct gestures. In a testing environment, with two sensors mounted on breadboards we have already been able to recognize 15 gestures.

VL053L0X Time of flight sensor (TOF) [5]
While the gesture information derived from accelerometer data is ideal for turning musical effects on and off, many DAW functions require values that vary continuously throughout a range. Adjusting the volume of individual samples within a piece, for example, would be difficult with only gesture data. The VL053L0X is another very small device (4mm x 3 mm x 1mm) that sends out light pulses and measures the time they take to reflect back to calculate digital distance measurements up to 2m. The Conductor will have a TOF sensor mounted on the palm, allowing the user to provide ranged data to the DAW by varying the height of their hand above a surface.

ESP32 [6]
The ESP32 is a microcontroller with built-in Wi-fi capabilities. It will live on the back of the user's hand where it will collect data from the sensors by controlling an I2C data bus. The ESP32 will compile and compress the data collected from each sensor, and filter out noise with a simple moving average. It formats the processed data for transmission and sends it via a small WiFi TCP server to the PC for gesture recognition. We have already developed and tested firmware to perform the ESP32's functions so we will be able to devote most of our remaining software development time to building the DAW plugin and user interface.

Neural Network [7-9]
A neural network is a form of machine learning that can 'learn' to approximate the mathematical relationship between a set of input data and a set of conclusions drawn from that data. To train the network we give it raw accelerometer data (the question) and a label corresponding to the known orientations of the sensor when the data was collected (the answer). The network then tries to guess the equation that links the two. It starts with a random equation and plugs the input into it. It gets a wrong answer that enters into a gradient descent algorithm, which tells it how to change the equation to reduce the difference between the true conclusion and the wrong result. As many accurately labelled input samples are run through the training algorithm many times, the equation gets better at predicting the correct conclusion. If there is a mathematical relationship between input and output the neural network will eventually learn to approximate it. It then becomes a fancy calculator, providing accurate predictions so long as the input is consistent with the training data.

We have already built a neural network and developed a training method that we have used to train 15 gestures using two sensors mounted on breadboards. Because

the relationship between sensor data and sensor orientation is relatively simple, the network is small, but still makes accurate predictions (~95%) in under 20ms, a tenth of the time it takes to transmit the data wirelessly.

Virtual Studio Technology (VST) [10]
Virtual Studio Technology type of software plugin to a DAW that extends artists' ability to modify, create, and perform their music. VSTs are designed to be highly modular and compatible with a wide variety of open source and commercial DAWs.

Figure 2. A user interface created with Reaper, a popular open source DAW [12].

We will develop a VST to bridge the gap between The Conductor and the virtual instruments in the DAW by translating the data from the neural network into MIDI commands that the DAW can make into sounds. JUCE is an open source C++ framework that simplifies the creation of VSTs with software libraries for designing graphical user interfaces and digital signal processing algorithms. JUCE has a large and active community of users and developers, and documentation and support is widely available. It will shorten our development time and allow users to easily incorporate our VST in their favourite DAW, providing a large potential user base for The Conductor.

Cooperative Management

Following our team's values of integrity, creativity and cooperation, our work will be organised cooperatively. Each team member is empowered to contribute ideas and labour to any aspect of the project, and our decision making and planning will be driven by collegial consensus. However, to ensure an equitable division of labour, work will be organised into sub-projects co-led by two team members each, while each team member will take co-leadership of two sub-projects. Co-leaders will be responsible for deliverables and timelines related to their sub-projects.

Hardware

Co-leads: Temo Guerrero, Italo Hernandez

- ☐ Build a wearable device with an embedded microcontroller, printed circuit boards, enclosures, motion sensors, and WiFi transceiver.

Gathering Data

Co-leads: Jose Figueroa, Joel Legassie

- ☐ Write software to recognize human gestures based on data gathered by the wearable device.

Applying Data

Co-leads: Jose Figueroa, Italo Hernandez

- ☐ Write software to translate gesture data into digital musical signals.
- ☐ Build a user interface to allow digital musical signals to interact with digital musical tools.

Documentation

Co-leads: Temo Guerrero, Joel Legassie

- ☐ Prepare promotional materials, design documents and other documentation (a website and user manual, or *Circuit Cellar* article).

Schedule

We will develop the prototype of The Conductor in three phases.

Proof of Concept and Circuit Board Prototype Design (Sept 2023)

- ☐ Develop firmware for the ESP32 to collect sensor data and send it to the PC client.
- ☐ Develop the neural network on the PC client, train it with 15 gestures and predict gestures in real time.
- ☐ Design and fabricate printed circuit boards.
- ☐ Plan and design VST software.

- ☐ Prepare documentation and marketing materials.

Prototype Construction (Oct 2023)

- ☐ Assemble components on printed circuit boards and test circuit boards.
- ☐ Design and print enclosures for circuit boards and mount them on a user's hand.
- ☐ Train neural network with data from the hand mounted prototype.
- ☐ Predict gestures with the hand mounted prototype.
- ☐ Prepare documentation and marketing materials.
- ☐ Design a second version of the circuit boards (time permitting)

Testing and Rework (Nov 2023)

- ☐ Print and test improved enclosures based on testing feedback.
- ☐ Assemble and test version 2 circuit boards.
- ☐ Rework circuit boards as required.
- ☐ Prepare documentation and marketing materials.
- ☐ Final presentation and symposium demonstrations.

Technology by Humans for Humans

We use a simple neural network to efficiently pick gestures out of motion sensor data. While the program does display some intelligence in 'learning' to correctly classify gestures, it does not have any more free will than a quadratic equation. Our goal is not to unlock and exploit the secret cheat codes of musical creativity, but to give creators tools and freedom to express themselves in new ways.

Financial

The Conductor will be developed with a materials budget of $250. The microcontroller will cost up to $30, while producing a prototype circuit board may cost up to $60. Accelerometers cost $2 each, and a laser time of flight sensor costs about $20 The 3-D printed enclosure is estimated to cost less than $5. This gives us a total BOM cost estimate of $134.54, (See Appendix A).

Research and Development costs will be the largest expense of the project. At 25 hours per week, for four team members over 14 weeks, the project will consume 1,400 hours of labour. At $50 per hour that translates to $70,000 of R&D labour.

The MiMu Glove is a similar product that retails for roughly $1500 [13]. With the above cost estimate, we would be able to break even at that price with the sale of only 53 units. However as an open source device accompanied by an educational service, The Conductor will be marketed with an alternative pricing model to be determined.

Conclusion

Music touches us in those unique and mysterious places that make us human. Modern digital technologies have given us great powers to manipulate the raw materials of sound, but they come with a price tag. As technology becomes more complex it reaches deeper into our personal lives. It requires us to bend our thinking and behaviour to its needs. Often, in a rush to make our machines more like humans, we force humans to become more like machines.

We believe this doesn't have to be the case. If it is done right, digital technology can be used to increase and enhance human creativity and cooperation, not hinder it. The Conductor will free electronic music artists from their screens, allowing them to communicate with audiences using emotive gestures instead of just turning knobs on a computer display. With the cost of materials amounting to $134.54, the conductor will give artists a low barrier to begin exploring a hands-on way to make electronic music for humans.

References

[1] W. Carlos, *Switched-On Bach*, 1968. Columbia Records, New York, NY.

[2] “Pause Giant AI Experiments: An Open Letter,” Future of Life Institute, [Online]. Available: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ [Accessed: Sept 11, 2023].

[3] G. Reid, "Moog Minimoog Model D," *Sound on Sound*, August 2016. [Online]. Available: https://www.soundonsound.com/reviews/moog-minimoog-model-d. [Accessed: Sept 11, 2023].

[4] "Monolithic and Wafer Level Packaged," MEMSIC. April 2015. [Online]. Available: https://mm.digikey.com/Volume0/opasdata/d220001/medias/docus/323/MXC40oxXC_Rev.B_4-24-15.pdf. [Accessed: Sept 11, 2023].

[5] "Adafruit VL53L1X Time of Flight Distance Sensor," Adafruit, Oct 2021. [Online]. Available: https://learn.adafruit.com/adafruit-vl53l1x/overview. [Accessed: Sept 11, 2023].

[6] “Introducing the new TinyS3 from Unexpected Maker” Unexpected Maker, [Online]. https://esp32s3.com/tinys3.html. [Accessed: Sept 11, 2023].

[7] H. Kinsley, D. Kukieła. *Neural Networks from Scratch in Python*. Harrison Kinsley; 2020.

[8] B. Krose, P. van der Smagt, *An Introduction to Neural Networks*. English Edition, November. 1996.

[9] R. Shamalik, S. Koli, “Real Time Human Gesture Recognition: Methods, Datasets and Strategies,” in *Recent Trends in Intensive Computing*, 2021.

[10] "Our Technologies," Steinberg. [Online]. Available: https://www.steinberg.net/technology/. [Accessed: Sept 11, 2023].

[11] "Features," JUCE. [Online]. Available: https://juce.com/juce/features/. [Accessed: Sept 11, 2023].

[12] J. Lendino, “Cockos Reaper Review: Flexible, Powerful Audio Editing for Less,” *PCMag*. [Online]. Available. https://www.pcmag.com/reviews/cockos-reaper. [Accessed: Sept 11, 2023].

[13] “MIMU | Home” [Online]. Available: https://mimugloves.com/. [Accessed: Sept 11, 2023].

Appendix A: Tentative BOM

Name	Manufacturer	Supplier	Quantity	Unit Cost	Total
ESP32-TinyS3 MCU	Unexpected Maker	DigiKey	1	$30.67	$30.67
I2C Multiplexer	Adafruit	DigiKey	1	$10.00	$10.00
Battery		Amazon.ca	1	$11.94	$11.94
Accelerometer	Memsic	DigiKey	4	$2.09	$8.36
OLED Screen	Adafruit	DigiKey	1	$17.98	$17.98
VL53L0X ToF sensor	Adafruit	DigiKey	1	$21.51	$21.51
Power regulator	ROHM Semiconductor	Mouser	1	$1.10	$1.10
3D printer Filament (PETG)		filaments.ca	100g	$29.95/kg	$2.95
Circuit Boards	JCLPCB	JCLPCB	2	$15.00	$30.00
				Total	$134.54

Note: The circuit board quote includes five copies of two separate circuit board designs and assembly of the surface mount accelerometers. The cost per unit produced will be considerably lower than this value.

Appendix B: Glossary

a.i
artificial intelligence algorithms that are able to mathematically 'learn' to associate properly formatted and accurate inputs with correct outputs with a greater (or lesser) degree of accuracy. A neural network is one form of a.i. It is not to be confused with true Artificial Intelligence which does not exist, yet.

C++ Framework
A C++ framework is a set of supporting software that provides extended functionality to the C++ applications.

DAW
A Digital Audio Workstation (abbreviated as DAW) is a software application that allows users to produce, edit, record, and perform music.

I2C
Inter-Integrated Circuit (I2C) is a serial communication protocol developed in 1982 by Philips Semiconductors used for connecting microcontrollers with peripherals such as sensors.

TCP Client
TCP (Transmission Control Protocol) is an internet protocol used to transmit information across the internet.

VST
Virtual Studio Technology is audio plug-in software used for connecting digital audio software with digital audio workstations.