
Titel

Project Report
Group MTA 16440

Aalborg University
Media Technology
Rendsburggade 14
DK-9000 Aalborg



Media Technology
Rendsburggade 14
DK-9000 Aalborg

AALBORG UNIVERSITY STUDENT REPORT

Title:

Title

Abstract:

Abstract

Theme:

Sound Computing and Sensor Technology

Project Period:

Spring Semester 2016

Project Group:

MTA 16440

Participant(s):

Alex Bo Mikkelsen
Allan Schjørring
Daniel Agerholm Johansen
Didzis Gailitis
Liv Arleth
Sebastian Laczek Nielsen

Supervisor(s):

Olivier Lartillot

Copies: ??**Number of Pages:** 82**Date of Completion:**

May 26, 2016

The content of this report is freely available, but publication (with reference) may only be pursued due to agreement with the author.

Contents

Preface	vii
1 Introduction	1
1.1 Initial Problem Statement	1
1.1.1 Motivation	1
1.1.2 Statement	1
1.1.3 Research Questions	1
1.1.4 Target Group	1
2 Problem Analysis	3
2.1 Research	3
2.1.1 Effects	3
2.1.2 State of the Art	4
2.1.3 Gestures	8
2.1.4 Conclusion	9
2.2 Problem Statement	10
2.3 Minimum Implementation	10
3 Design	11
3.1 Concept	11
3.2 Storyboard	11
3.3 Sketching and Testing	12
3.4 Lo-Fi	15
3.4.1 Affordance Scheme	15
3.4.2 The Test	16
4 Implemented Theory	17
5 Implementation	19
5.1 Theory	19
5.2 Hardware and Software	19
5.2.1 Hardware	19
5.2.2 Schematic	19
5.2.3 Software	22

6 Evaluation	25
7 Discussion	27
8 Conclusion	29
Bibliography	31
A Appendix	33

Preface

Aalborg University, May 26, 2016

Alex Bo Mikkelsen
<amikke13@student.aau.dk>

Allan Schjørring
<aschja14@student.aau.dk>

Daniel Agerholm Johansen
<djohan14@student.aau.dk>

Didzis Gailitis
<dgaili14@student.aau.dk>

Liv Arleth
<larlet14@student.aau.dk>

Sebastian Laczek Nielsen
<slni14@student.aau.dk>

1. Introduction

1.1 Initial Problem Statement

1.1.1 Motivation

It is possible to use voice effects while performing. Many useful effects for performing exist, and it is possible to change the parameters of these effects to one's liking in real time. A problem can be applying an effect and/or changing the effect parameters while performing without having a big control board in front of the performer which may cause a disruption of the performance.

1.1.2 Statement

How does one create a system that applies voice effects to a voice in real-time?

1.1.3 Research Questions

- What are the most common voice effects?
 - Which of these effects would singers have the desire to change during a performance?
- Does any existing technology use body gestures or sensors to apply effects?
- Which gesture should be used with which effect, and how?

1.1.4 Target Group

The criteria for the target group in this project are:

- Should be able to sing
- Should know the basic theory of singing
- Preferably should know about voice effects

The target group consists of singers that know about voice effects. They should not play an instrument while singing because they must be able to use their body

for controlling the effects. There is no specific genre or type of singer as the only criteria is that they know the technicalities behind singing.

People who fulfil these criteria could for example be solo singers, choirs and band singers.

2. Problem Analysis

This chapter will strive to answer the research questions posed in 1.1.3 and any other relevant information that arises. Additionally, a final problem will be stated after the research is complete.

2.1 Research

This section focuses on the research necessary for the basic understanding of the subject. This will include a short description of some effects that could be applied by the system, the current state of the art within the field, and some theory behind the use of gestures.

2.1.1 Effects

Many voice effects exist today. Some effects are used in most music, and some are not. The effects can be really subtle, or really noticeable. In this section, there will be short descriptions of some common effects.

Delay

A delay effect creates a repetition of the original sound after a period of time[1]. By using the delay effect, it is possible to simulate the sound of the echo created when yelling into a cave or over a canyon.

Reverberation

When sound reflects off surfaces in a confined space, its called natural reverberation[2]. Reverberation like this works best when the sound hits hard surfaces. For example, the sound effect that comes when you sing or yell in a church, is reverberation. The sound bounces all around the church's hard walls. Digitally, the way to simulate reverberation is to use a multitude of delays and feedback. This then creates a series of echoes that then slowly decays.

Pitch Shift

The frequency of a harmonic sound is called its pitch[3]. By shifting the pitch, the sound will effectively become deeper or higher. An example of this is the voice that anonymous people get when they want to hide their voice, this is a lowered pitch. Another example is the “chipmunk voice”, which is achieved through a raised pitch. Pitch shifting can be done by using the "phase vocoder", which is a digital signal processing technique[4]. The phase vocoder works by analysis and synthesis. The analysis part takes the signal, and models it as a sine wave in which one can find the amplitude, phase, and frequency of the sine wave. In the synthesis part, one can manipulate these parameters. The phase vocoder can do many things, e.g. change the pitch of a sound without changing the duration of the sound - make a sound deeper or higher.

Pitch shifting is also used to create the harmonizer effect. It takes the input voice and shifts its pitch a certain amount, and then adds it as an additional voice. This can effectively simulate a choir.

Auto-Tune

The Auto-tune effect corrects a singer’s voice to the correct tone[5]. This can be really subtle or plainly obvious. The user just needs to choose a reference of scales or tones and the amount of correction that needs to be made, and then the Auto-tune will make the proper adjustments.

Vocoder

The Vocoder effect combines a singer’s voice with another sound - that could be the sound from an instrument or a synthesizer[6]. The effect can make the voice sound like a robot. The vocoder needs two inputs, the voice and e.g. an instrument. The fundamental frequencies of the voice are converted to levels of amplitude on a series of band pass filters, which then are passed through the instrument sound.

Discussion

2.1.2 State of the Art

To gain understanding of what is possible, a study of the state of the art was conducted with the focus on commercial artefacts used for real-time alterations.

TC Helicon Perform V

The TC Helicon - Perform V is a vocal multi-effects processor that attaches to a microphone stand, as seen in figure 2.1[7]. It has three effect buttons, three preset buttons, a big knob, and other buttons. The effects are reverb, echo, “double” (harmonizer), equalizer, compressor, and many more. It is possible to download an

2.1. Research

application that can connect with the Perform V. The application has many pre-made sounds, and it has a wireless connection.



Figure 2.1: TC Helicon Perform V[7]

The Perform V is good for live performing if the singer has the processor in front of them, on the microphone stand. Preset buttons make it easy to change effect quickly. If the singer plays an instrument, it is probably difficult to change effects without interrupting the instrument playing. Another downside is that the singer is limited to only three presets, and only one knob to turn.

Electro Harmonix Voice Box

The Electro Harmonix Voice Box is a more advanced processor than the TC Helicon[8]. It has six knobs: blend, two reverb knobs, “gender bender”, voice mix, and “Mode”, as seen in figure 2.2. It has nine different modes, which includes different kinds of harmonies, unison-whistle, and a vocoder, which the TC Helicon does not have.



Figure 2.2: Electro Harmonix Voice Box[8]

The Voice Box has to be on a flat surface, like the floor or a table and is most often used as a pedal. It is possible to insert an instrument into the pedal, so it can be used for the vocoder. The Voice Box has many effects and knobs - this can make changing effects and effect parameters difficult, even more if the pedal is on the floor.

Mi.Mu Gloves

The Mi.Mu Gloves are gloves made for making music, and controlling sound[9]. They are made by scientists, musicians, and artists, and have been in development since 2010. They are wearable, and can be used by one or both hands, see figure 2.3. The gloves have been through many iterations, and they are open source. The gloves use gestures, hand and finger movement, finger placement, and other features to control sounds and effects. The hardware includes an ArduImu, flex/bend sensors, accelerometer, gyroscope, haptic motors, LED's, Wi-Fi compatibility, and provides other capabilities.

2.1. Research



Figure 2.3: Mi.Mu Gloves[9]

The gloves are bluetooth or Wi-Fi connected, so the person using the gloves are free to move around, and does not have to worry about wires. They are also battery powered. Since the gloves are open source, you can make your own - many different gloves exist - some are simple, and some are complex.

HandySinger: Expressive Singing Voice Morphing using Personified Hand-puppet Interface

Yonezawa et al. made a glove that controls voice effects [10]. The wearer of the glove controls a puppet and makes hand gestures, see figure 2.4.



Figure 2.4: HandySinger Glove [10]

They believe that using a puppet interface will increase the expressiveness of the

user's singing voice. The glove itself has seven bend sensors, and two pressure sensors, see figure 2.5.

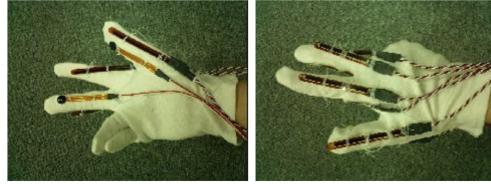


Figure 2.5: HandySinger Glove sensors [10]

The glove measures both forward bend and backwards bend. The gestures that the users can make are: bend back clasp, drooping, stretching, and bend back. The parameters that the gestures change are: "dark", "whisper", "wet", and volume. Yonezawa et al. found that users with small hands had trouble using the glove effectively. Nevertheless, they confirmed it was easy to gesture with the hand-puppet and that the gestures reflect the voice expression changes.

The 'One-Person Choir': A Multidisciplinary Approach to the Development of an Embodied Human-Computer Interface

The study by, Maes et al. [11] utilises body gesture to enhance a singer's voice. The system is a human-computer interface that use gestural control for harmonising a singing voice. The system is operating in real-time, which means it is possible to use it during live performances. The system uses pre-configured models to control the harmonisation, and the singer can eventually use this to enhance his or her singing voice. During their research, they found that gesture control is a big part of singing, which also helps the perception of the singing. The movement of the upper body is the primary gesture used in the system which means that the singer has sensors attached to the upper body.

2.1.3 Gestures

When designing a way to control effects, there are several ways to approach it. One of these ways is through gesture control.

There are also several ways to approach gesture control. For example, there are three stages of a gesture: registration, continuation, and termination[12, pp. 127-134].

2.1. Research

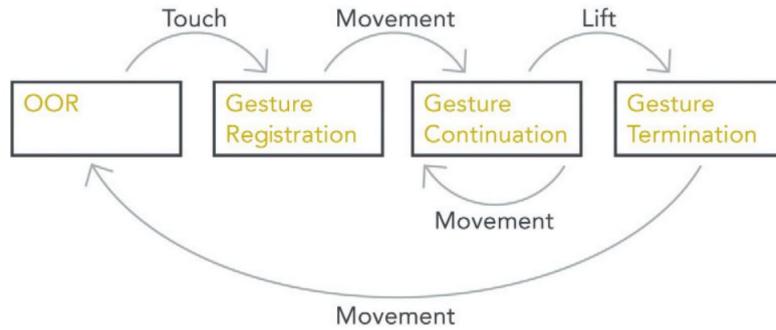


Figure 2.6: Out of Range(OOR) and the three stages of gestural input [12, pp. 127]

The registration stage is when the system registers that the user would want to perform a gestural input, e.g. when you place your finger(s) on a touch display.

The next stage is continuation. In this stage, the user uses movement to adjust the parameters of the gesture, e.g. when you move two fingers away from each other in order to zoom, on a touch display.

The final stage is termination, where the user simply ends the gesture, e.g. by lifting their finger(s) from a touch display.

In some cases the continuation stage can be skipped. Whether or not it is a good idea to skip this stage is decided by the number of different gestures one wants to implement, and whether they want the gesture to be able to change the parameters or not, e.g. a zoom that does not have one set amount of zoom, but rather it is specified by how much you move your fingers. By removing the continuation stage, it also removes the possibility to differentiate between a lot of gestures.

A good thing to do, when choosing a gesture for a specific action, is to keep it as unambiguous as possible. This helps reduce future errors. This being said, it is also important to minimise the amount of steps the user has to go through for the gesture.

While the previously mentioned ‘zoom’ gesture is a rather good example of a gesture, a bad example of a gesture would be the use of the ‘flick’ action to execute a gesture. The ‘flick’ action is when you execute an action by ‘flicking’ an object to e.g. delete it. The bad thing about this design is the fact that you have to specify a border between the action of moving something and ‘flicking’ something

2.1.4 Conclusion

There are numerous sound effects each unique in the way that it works and is applied, so understanding them is important as each of them has different limitations or applications in a real-time setting. By going through each of them and learning about them, the effects most relevant to the purpose of the project was picked and described.

From the state of the art it can be seen that extensive effort has been put into creating many different devices of varying complexity and limitations. In design most of them vary between a glove and a plug in device, those design differences also provide different accessibility while performing. Additionally, the state of the art provided information about which voice effects professional singers would usually like to adjust and how.

Finally, by researching what a proper gestural input should consist of and how it should be performed, it was discovered that there are several stages to a gesture, and some of the stages are more important than others, depending on the purpose of your gesture. Furthermore, when looking at the previous state of the art examples, it can be concluded that two of the most common gestures for a singer to apply an effect and change its parameters, is by pushing a button and to turn a knob, as can be seen in figures 2.1 and 2.2.

2.2 Problem Statement

Based on the findings from the research a final problem statement has been made:

Audio effects for the voice exist but they are impractical to change while performing. Technology exists that address this problem, e.g. gloves that use sensors. We want to make a glove that can change voice effect and their parameters.

Success Criteria

- The system should have at least two effects
- Users use the right gestures to change the effects
- The system does the right action to the gesture - does not misinterpret

2.3 Minimum Implementation

- The system must implement the use of an Arduino
- The system must implement the use of sensors applicable to the Arduino
- The system must implement audio processing
- The system must be implemented using the Arduino software and PD Extended
- The system must get audio from a microphone

3. Design

This chapter describes the design process for this project. Initially, the concept of the product is presented...

short introduction to what will be addressed in this chapter

3.1 Concept

The concept of the product is to be able to apply voice effects in real-time without having to turn to a panel or having someone do it for you.

A thing most singers almost always have available are their hands. Therefore a device controlled by the hands movements seems the obvious choice.

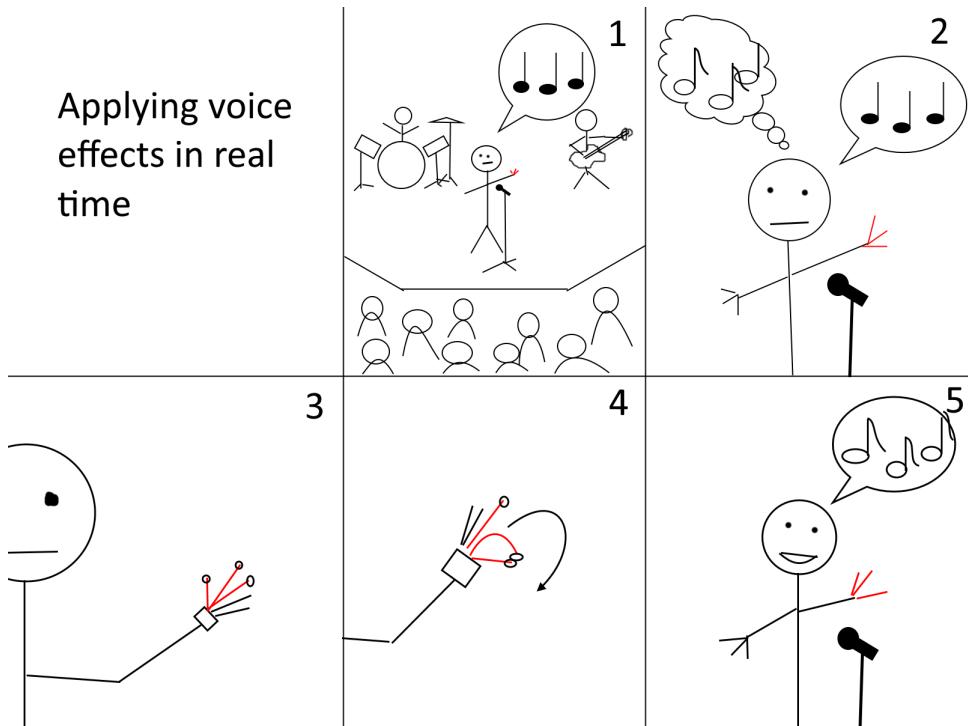
The device will implement a gyroscope to sense the movements of the hand.

The device then needs to be told that an effect has been initiated. This is done by connecting the thumb to the finger in control of the desired effect.

- Harmonising: This will be controlled by turning the hand while having thumb and a finger pressed together, like turning a knob or volume control.
- Pitch: This will be controlled by lifting or lowering the hand while having thumb and a finger pressed together, like pulling a slider up or down.

3.2 Storyboard

To understand the concept and how it would work in a real scenario a storyboard was created to explain how, what and where to use the product, see figure 3.1. this storyboard was part of the proof of concept that together with the initial sketches would explain the idea better.

**Figure 3.1:** The Storyboard

In the first frame a band is shown playing a concert on a stage, with the lead singer singing. The second frame shows the lead singer singing, but wanting to sound different. In the third frame the device is shown on the singer's hand. Fourth frame illustrates the singer performing a gesture to apply the desired effect. The final frame then shows the singer sounding how they wanted to sound.

3.3 Sketching and Testing

The initial sketches and the first concrete design of the device have copper foil on thumb, index finger and middle finger, see figure 3.2. When pressing the index or middle finger together with the thumb, a connection is made in the system that activates the assigned effect.

3.3. Sketching and Testing

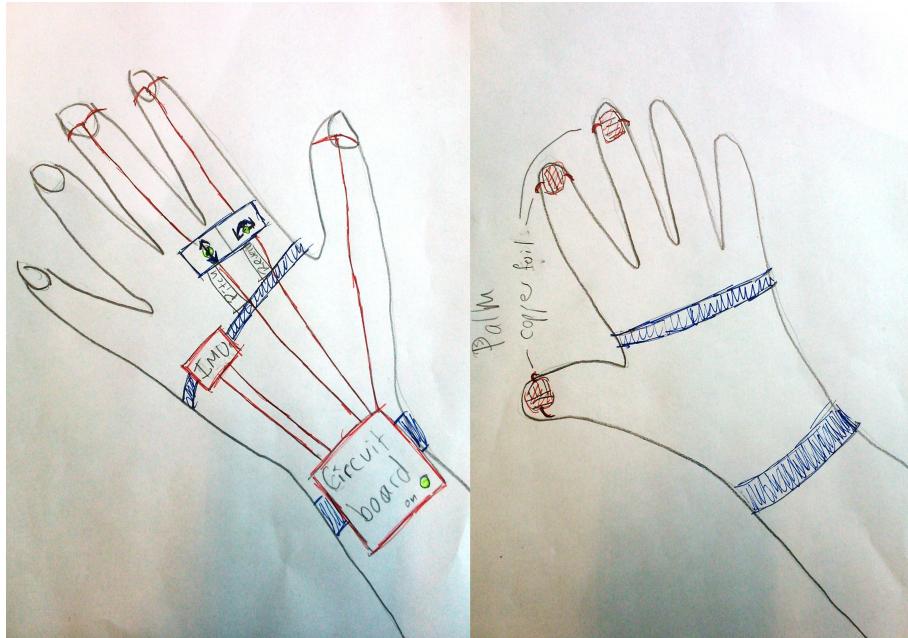


Figure 3.2: Two sketches showing idea of the system

On the knuckles there are illustrations of the gestures, that you are supposed to do to manipulate the effects. Beneath those are small labels with the name of the effect. In this design reverb is used instead of harmonising. This was later changed, since reverb is not manipulated quite as much as harmonising.

The sensor is attached to the hand by a velcro strip, as is the circuit board.

A quick informal test with three participants was conducted and they were told what the drawing was supposed to be and what it should do. They then had to figure out based on the sketch how to do those things. From this test, a few key things were learned:

- They all had difficulty figuring out how to get to the activate stage. None of them connected their fingers.
- Most eventually figured out which type of gesture in general had to be done, but only after some trial and error
- They all found out which finger created which effect

Based on that the next focus was on creating some feedforward and perceived affordance that tells the user, that to activate the glove you need to connect two fingers.

The second sketch changed the illustrations since people had a hard time immediately recognising the correct gestures with the old ones, as seen in figure 3.3.



Figure 3.3: Second set of sketches after some changes

Colour was also added to the copper foil, a different one on the index and middle finger and then both on the thumb. This was done to create perceived affordance between the fingers and thumb.

LEDs were added on the circuit board to create some feedback for the actions. The middle LED showing if the system was on or off. The two other LEDs showing an increase or decrease in the effect, with a minus and a plus sign.

A quick informal test was done with two participants with the revised sketch. Now there was a better indication that they needed to connect two fingers, but not anything that indicated that they needed to stay connected. Additional results from the test:

- One suggested that instead of the on/off LED, maybe a connected/not connected LED.
- The arrows were found to be confusing for one tester.
- Another tester easily understood the pitch action but was a bit confused with the placement of the arrow on the harmonise action.
- One tester thought that the dual colour on the thumb suggested that both actions could be done at the same time.

3.4. Lo-Fi

- The plus and minus LEDs confused one tester, but this could also be because the drawing was unclear.

3.4 Lo-Fi

Based on the previous test, a new design iteration of the glove was made, now also in the form of a lo-fi model, as seen in figure 3.4. This lo-fi was created in order to make a mental model test. The desire was to both get some feedback on the general design, but also to weigh the users' mental model against an affordance scheme.

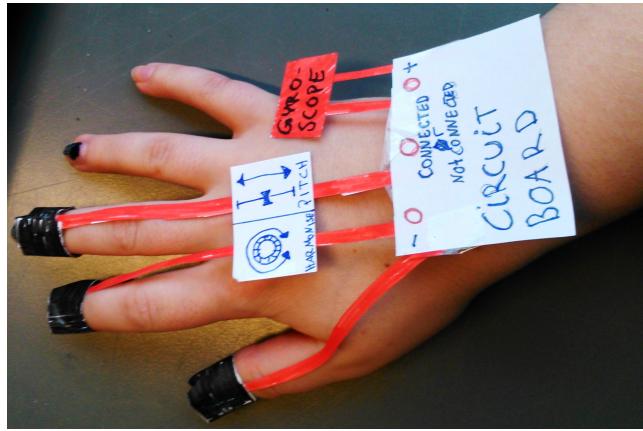


Figure 3.4: The Lo-fi of the system

3.4.1 Affordance Scheme

The affordance scheme is separated into three categories: perceived affordance, feed-forward, and feedback. Perceived affordance is the perception that something is interactable, e.g. a user would assume a button can be pressed no matter what state the system is in. Feedforward is what is expected to happen after a certain action, e.g. after pressing the "On" button, the system will turn On. Finally, feedback is what the system does to indicate an action has taken place, e.g. after pressing the "On" button, a display lights up and says "Turning On".

The following table shows the affordance scheme for the current lo-fi design, seen in figure 3.4.

State	Perceived Affordance	Feedforward
Inactive	It is a wearable glove	
Active		Connect fingers + perform gesture on labels = effect c
Performing Gesture		Connect fingers + perform gesture = effect chang

3.4.2 The Test

The test was initiated with a short introduction for the user, without revealing too much. Then the participants were asked to explain everything they saw and assumed about the glove, and to try to put it on. Following this, they were given the tasks of turning harmonisation up and pitch down. Finally, some follow-up questions were asked to figure out why they did what they did.

Seven people participated in the test. Here are some results from the test:

- When trying to apply the effects, 6 out of 7 testers' first reaction was to not connect their fingers
- 5 out of 7 did the correct gestures, based on the illustrations on the knuckles
 - 1 of the two who did not do the correct gestures, did what she did because she thought the illustrations were interactable
- 6 out of 7 found the illustrations helpful
- All participants figured out that the device was a glove to wear on the hand
 - 5 out of 7 put it on correctly, as seen in figure 3.4
- All participants were in doubt of what the LEDs were for

The table below shows a revised affordance scheme based on the user feedback. The parts of the system that the participants had a hard time understanding are shown in italic writing in the table. Additionally, the parts which were generally understood by the users has been underlined. The participants had a hard time understanding some of the feedback in the system, particularly the LEDs. While they knew that they were LEDs, they were not sure what each of them was signifying.

State	Perceived Affordance	Feedforward
Inactive	It is a wearable glove	
Active		Connect fingers + perform gesture on labels =
Performing Gesture		Connect fingers + perform gesture = effect

4. Implemented Theory

5. Implementation

The following chapter describes how the prototype has been implemented and the materials needed to build the prototype.

5.1 Theory

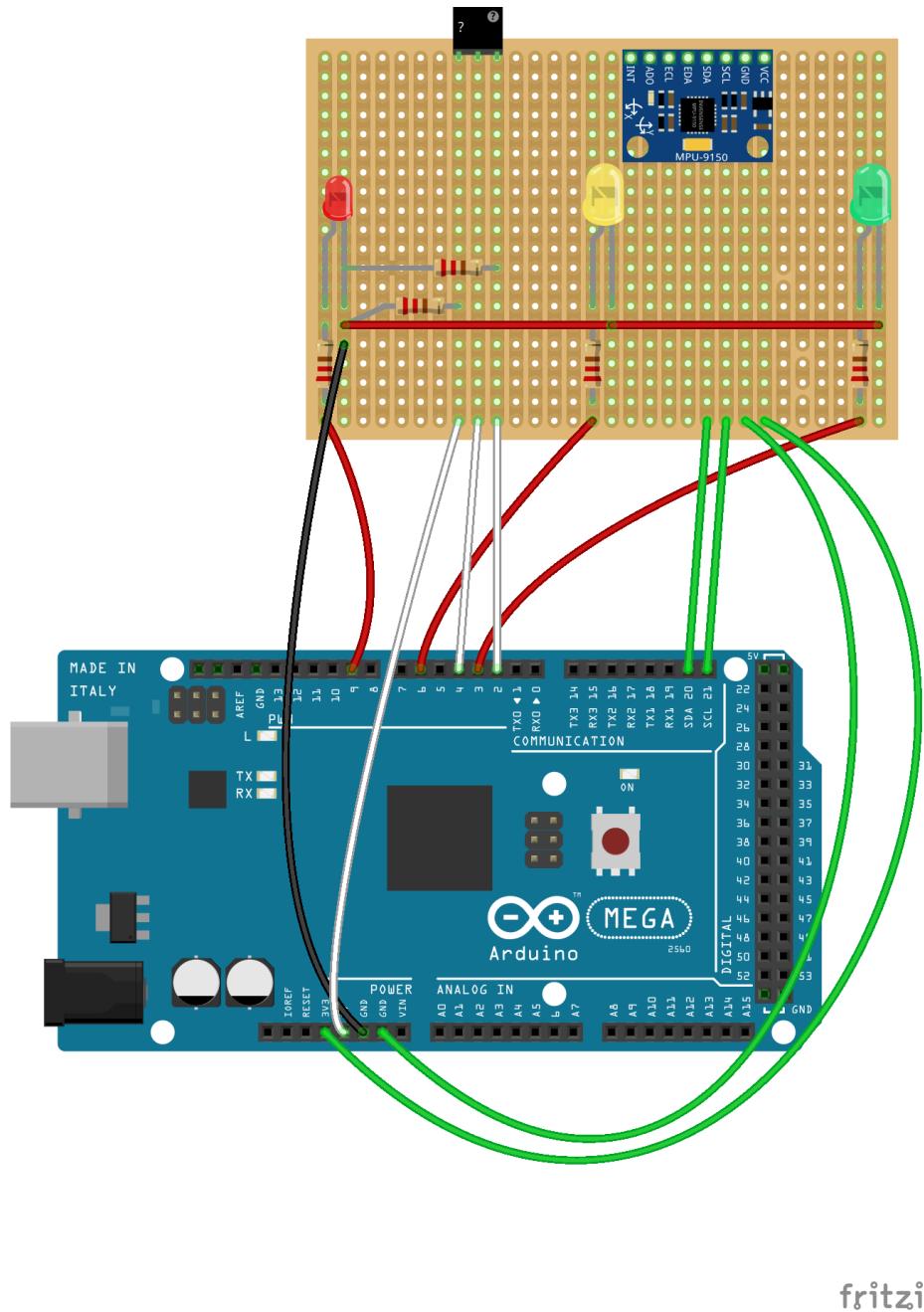
5.2 Hardware and Software

5.2.1 Hardware

The prototype consists of an Arduino Mega 2560[13], with a connection to a circuit board with an inertial measurement unit(IMU). The IMU uses the MPU 9150 sensor with 9 degrees of freedom[14]. It has a tri-axis gyroscope, magnetometer, and accelerometer.

5.2.2 Schematic

The following figure 5.1 shows the connection from the Arduino to the circuit board with the IMU. The black switch at the top of the circuit board represents the three connections, which is put on the thumb, index and middle finger. These are velcro based, with copper tape on them, to create the connection when pressed together.

**Figure 5.1:** Prototype Schematic

The circuit board consists of three LEDs, a yellow, green and red. They light up at different points, which is explained in the next section of this chapter. The LEDs are connected to resistors of 220 ohm which is wired to the arduino as seen in the schematic. The green LED is connected to pin 3. The yellow LED is connected to pin 6 and the red is connected to pin 9. This is essential to make it work in

5.2. Hardware and Software

the Arduino code, which will be described later. The black wire connects all the LEDs and runs from them, to the ground on the Arduino. The blue circuit board represents the IMU. The necessary connections are: SDA, SCL, GND and VCC. The SDA connection is connected to pin 20 and the SCL is connected to pin 21. GND is connected to ground on the Arduino and the VCC connection powers the sensor and is therefore connected to 3.3V on the Arduino. The white wires are also connected to the Arduino. The first wire from the left, which goes to the thumb, is connected to the 5V on the Arduino. The white wire beside it, is connected to the index finger, and pin four on the Arduino. The last white wire is connected to the middle finger and pin two on the Arduino. The wire which is connected to the thumb creates a circuit to the LEDs whenever the connection is created between either the thumb and index finger or thumb and middle finger.

The LEDs has to be connected to a pull-down resistor[15], which makes sure, that the Arduino does not 'float' between two different values. A pull-down resistor ensures, that the value is zero, when no active device is connected. The following figure ?? shows what a circuit with a pull-down resistor might look like.

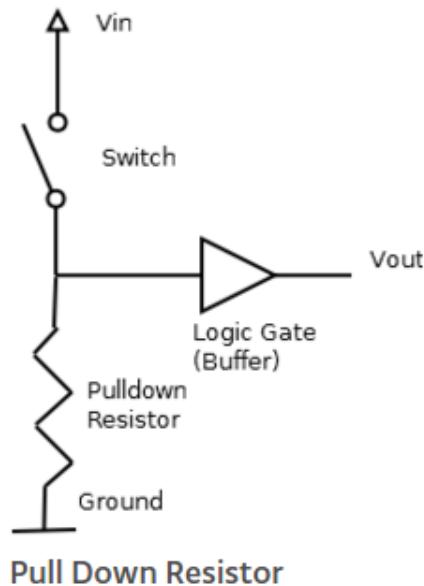


Figure 5.2: Schematic of Pull-down resistor[?]

The following figure shows the prototype. As one can see, the velcro for the fingers has copper tape on it, which has been glued onto the velcro. This means that whenever the copper tape gets connected with the copper tape on the thumb, it creates the connection as mentioned earlier. The prototype has cloth sewed onto it, which makes it possible to wrap it around the wrist.

5.2.3 Software

The following section describes the software part of the implementation.

Arduino

The Arduino language is based on C/C++, and whenever a sketch is compiled, it is sent to a C/C++ compiler [16].

PureData

The audio processing has been done in PureData(PD)[17], which is an open source programming language. It is used to generate and process sound, in a graphical way. The program uses patches where one can create objects which makes it possible to create different audio effects. The following subsection describes how PD has been used in this project, and explains the patches made for this project.

The following figure 5.3 shows the pitch shifting patch made in PD.

5.2. Hardware and Software

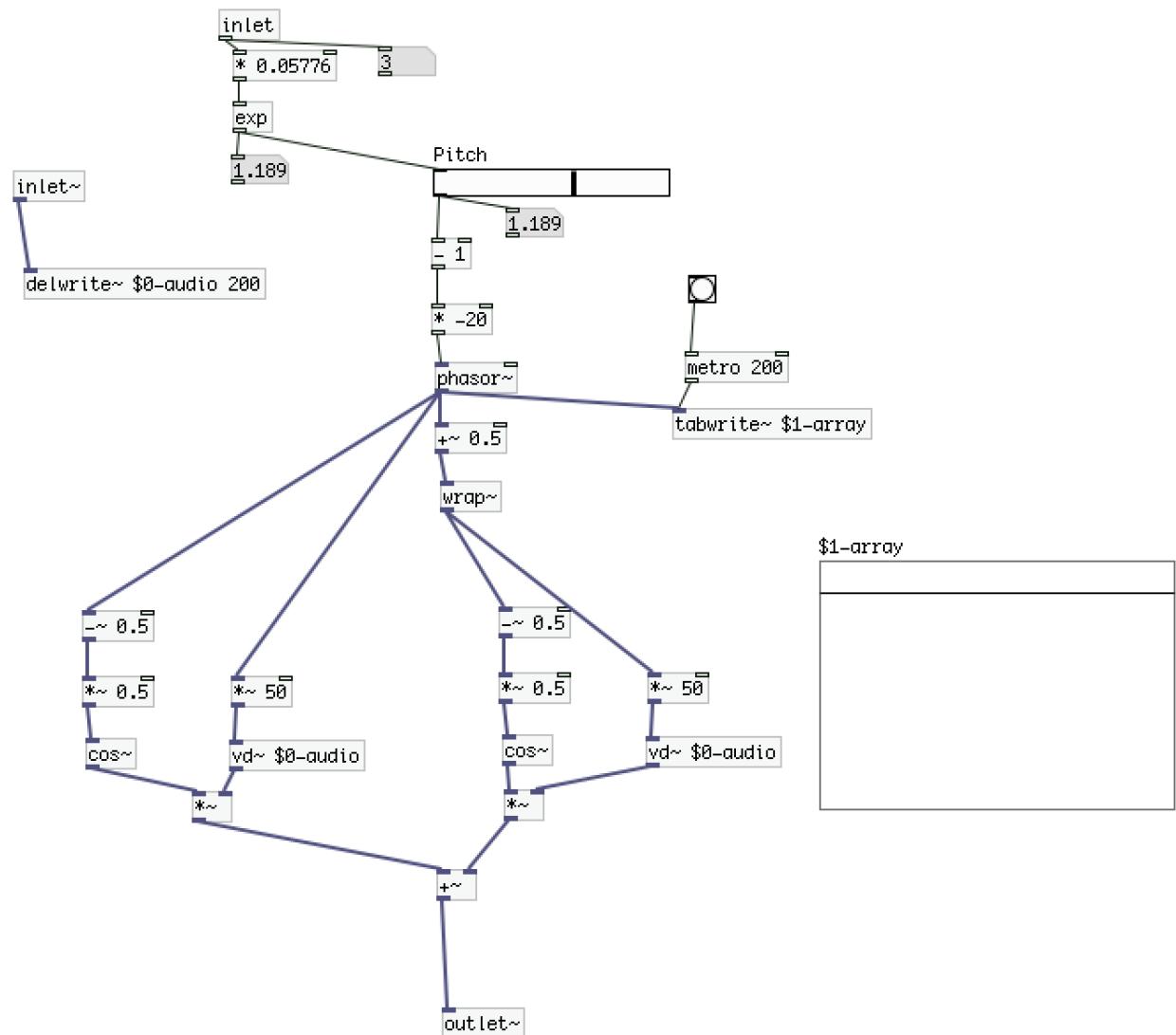


Figure 5.3: Pitchshifting in PureData[?]

6. Evaluation

7. Discussion

8. Conclusion

Bibliography

- [1] C. Loeffler, “A brief history of time-based effects (part 1).” <http://www.harmonycentral.com/articles/exploring-time-based-effects>, September 2014.
- [2] N. Redmon, “A bit about reverb.” <http://www.earlevel.com/main/1997/01/19/a-bit-about-reverb/>, January 1997.
- [3] Katjaas.nl, “Pitch shifting.” <http://www.katjaas.nl/pitchshift/pitchshift.html>.
- [4] M. Dolson, “The phase vocoder: A tutorial,” *Computer Music Journal*, vol. 10, no. 4, pp. 14–27, 1986.
- [5] A. Hadhazy, “What’s auto-tune and how does it work?” <http://www.livescience.com/11046-auto-tune-work.html>, September 2010.
- [6] Innovativesynthesis.com, “Introduction to vocoders.” <http://www.innovativesynthesis.com/introduction-to-vocoders/>.
- [7] TC-Helicon.com, “Perform v.” <http://www.tchelicon.com/en/products/perform-v/>.
- [8] EHX.com, “Vocal harmony machine/vocoder.” <http://www.ehx.com/products/voice-box>.
- [9] I. Heap, T. Mitchell, K. Snook, S. Madgwick, H. Perner-Wilson, A. Stark, R. Freire, and C. V. D. Berg, “Mi.mu gloves.” <http://mimugloves.com/>, 2015.
- [10] T. Yonezawa, N. Suzuki, K. Mase, and K. Kogure, “Handysinger: Expressive singing voice morphing using personified hand-puppet interface.” http://www.nime.org/proceedings/2005/nime2005_121.pdf, 2005.
- [11] P.-J. Maes, M. Leman, K. Kochman, M. Lesaffre, and M. Demey, “The one-person choir: A multidisciplinary approach to the development of an embodied human-computer interface.” http://www.mitpressjournals.org/doi/pdf/10.1162/COMJ_a_00054, 2011.

- [12] D. Wigdor and D. Wixon, “Brave nui world.” http://www.gm.fh-koeln.de/~hk/lehre/sgmci/ss2015/Literatur/Wigdor_Wixon_-_Brave_NUI_World.pdf, 2011.
- [13] “Arduino mega 2560.” <http://www.arduino.cc/en/Main/ArduinoBoardMega2560>.
- [14] Sparkfun, “Sparkfun 9 degrees of freedom.” <https://www.sparkfun.com/products/retired/11486>.
- [15] “Pull down resistor.” <http://playground.arduino.cc/CommonTopics/PullUpDownResistor>.
- [16] Arduino.cc, “Frequently asked questions.” <http://www.arduino.cc/en/main/faq>.
- [17] PureData.info, “Pure data.” <https://puredata.info/>.

A. Appendix