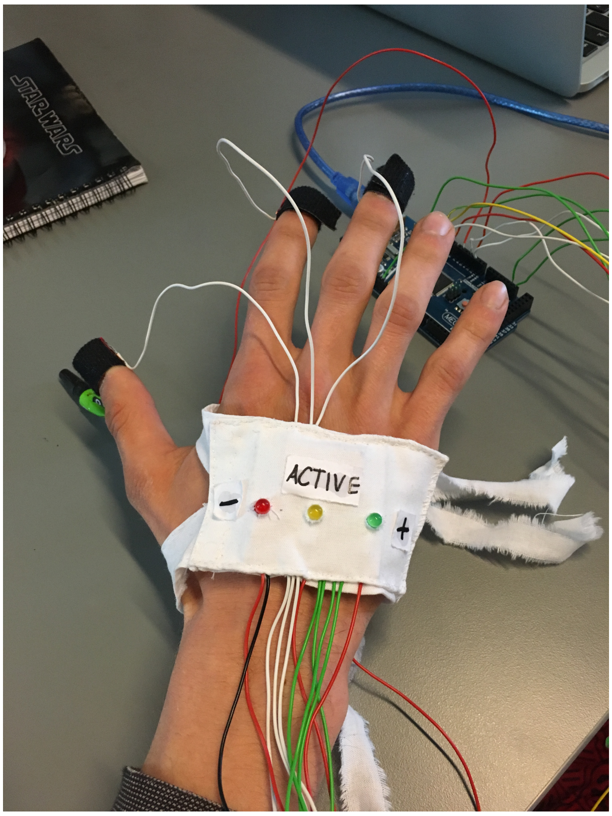
Real-Time Voice Manipulation Using a Wearable Device



Project Report

Group MTA 16440

Aalborg University
Media Technology
Rendsburggade 14
DK-9000 Aalborg

AALBORG UNIVERSITY
STUDENT REPORT

Media Technology
Rendsburggade 14
DK-9000 Aalborg

Title:
Real-Time Voice Manipulation Using a Wearable Device

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: 8

Number of Pages: 65

Date of Completion:
May 26, 2016

Abstract:

Today the use of voice effects are common in music. It can however be difficult to change these in real time during a performance. This problem is addressed in this project by focusing on creating a more intuitive and simple solution than the ones that already exist. In this project a wearable device, to put on the hand, is created to apply voice effects to a performance in real time. Different effects has been researched and pitch shift and harmonisation were eventually chosen to be implemented. This was done by using a gyroscope to track hand movement which is connected to an Arduino. The audio is processed in Pure Data. It was concluded through testing that the system was fully functional. It is not possible to draw a conclusion from the user test, since the target group was not optimal.

Contents

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

To use voice effects when performing has become more and more common through recent years. The voice effects grant the performers a wide range of opportunities. Each effect has its own unique property, and each has both pros and cons. An example of this is auto-tune.

The pros of auto-tune is that a singer can sound like they're singing in tune, pitch perfect and with perfect timing, but the downsides of this could be making everything too perfect, that way taking away from the live performance [1].

If the performer wants to use voice effects during their performance they must plan it all very carefully and have pre-sets they can activate or have another person control them during the performance. This removes a lot of control from the performer, which could be a hindrance for some.

This project focuses on making the effects readily available for the performer in real-time. This will be done by making a wearable device, which the performer can use to apply and change voice effects.

Initially, research will create understanding of what is needed to create a proper device a performer could use during a performance.

A glove will be made for the performer to apply the effects via gestures. The design will be focused around making the device intuitive and easy to use.

When implementing the prototype, the Arduino and Pure Data platforms will be used alongside an Arduino circuit board, to be able to communicate between the audio processing of Pure Data and the physical device via the Arduino.

Two voice effects will be implemented with appropriate gestures and parameters. While only two will be implemented at this stage, the system will be able to support even more effects.

The prototype will be tested for technical faults to be sure that it is functional, and it will be tested by users to learn whether the device is intuitive to use. This data will then be analysed both quantitatively and qualitatively to give a broader view of the results.

Based on the results from the evaluation, a discussion and conclusion will be made.

1.1 Initial Problem Statement

The problem addressed in this project deals with real-time access to voice effects for performers.

Some performers find it useful to use voice effects while performing. This project focuses on making the effects accessible and interchangeable in real-time. Instead of most commonly only being applied in set amounts by the push of a button, it could also be beneficial for the performer to be able to tinker with the effect parameters in real-time. In many cases it is not the performer that controls which effects are applied and when, rather it is applied by someone backstage or similar.

Additionally, many existing devices used for adding voice effects can be difficult to use for example during a concert. This could be a pedal, which often just has set effect parameters, where the performer pushes a button and e.g. a harmony is applied. The problem with the pedal is that it is stationary and the performer will have to reach for it to apply an effect, which might be inconvenient in some cases.

In this project, it is the intention to give the performer full control of the effects and its parameters while moving freely across the stage. It could be interesting to make a wearable device that accomplishes this, while also being intuitive for the performer to use.

1.1.1 Statement

In order to obtain the necessary information regarding the subject the following has been stated.

How does one create a wearable device that applies voice effects to a voice in real-time, while also having an intuitive interface design?

1.1.2 Target Group

The criteria for the target group in this project are:

- They should have experience with performing
- They should have a basic understanding of how voice effects works

The target group consists of performers who know about voice effects. They should not play an instrument or similar while performing since they must be able to use their body for controlling the effects. The target group has no specific genre or type of performer as a criterion, instead they only need to know the technicalities behind performing.

Prime examples of people who fulfil these criteria could be singers, stage actors or stand-up comedians.

2. Problem Analysis

This chapter focuses on the research necessary for basic understanding of the subject. The focus lies on answering the following research questions:

- What are the most common voice effects?
  - Which of these effects would performers 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?

The research will include a short description of effects that could be applied by a prototype, the current state of the art within the field, and theory behind the use of gestures.

2.1 Effects

Many voice effects exist today. Some effects are used by singers and some are used by other performers. The effects can be subtle or stand out. In this section there are short descriptions of some common effects.

2.1.1 Delay

A delay effect creates a repetition of the original sound after a period of time[2]. 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. Some people may find this desirable to use in their performance.

2.1.2 Reverberation

When sound reflects off surfaces in a confined space it is called natural reverberation[3] commonly referred to as reverb. Reverberation like this works best when the sound hits hard surfaces. For example, the sound effect that is heard when you sing or yell in a church, is reverberation.

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 decay.

If the amount of reverberation is small, it can be used to simulate singing in a small or medium sized room, even though the voice input has no natural reverberation. If the amount of reverberation is big, it can make the voice sound like singing you are in a church. This can make a person's voice sound "big".

2.1.3 Pitch Shift

The frequency of a harmonic sound is called its pitch[4]. By shifting the pitch, the sound will effectively become deeper or higher. An example of this is when a lowered pitch applies a deeper voice to anonymous people when they want to hide their voice. Another example is the "chipmunk voice" which is achieved through a raised pitch.

Pitch shifting can be done by using a phase vocoder which is a digital signal processing technique[5]. The phase vocoder works by analysis and synthesis, where the analysis 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 one can manipulate these parameters.

The phase vocoder has many uses, e.g. change the pitch of a sound without changing the duration of the sound which means it can make a sound deeper or higher in real-time.

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

2.1.4 Auto-Tune

The auto-tune effect corrects a singer's voice to the correct tone[6]. 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.

2.1.5 Vocoder

The vocoder effect combines a performer's voice with another sound which could be the sound from an instrument or a synthesizer[7].

The effect can make the voice sound like a robot.

The vocoder needs two inputs, the voice and an instrument. The fundamental frequencies of the voice are converted to levels of amplitude through a series of band pass filters, which are then passed through the secondary sound, which could be an instrument.

2.1.6 Discussion

Changing effect parameters can benefit some effects but not all. In a regular reverb effect one can change the amount of reverb by adding high or low-pass filters. Changing reverb parameters would probably be too subtle for the untrained to notice.

The delay effect has the parameters: amount of feedback, length of delay, and a filter. The amount of feedback and length of delay are parameters to consider, since it would be easy to hear a difference.

Pitch shifting is also a good effect for changing parameters. Here it would be to raising or lowering the pitch. The harmoniser uses multiple pitch shifters but the parameters one could change would be the amount of harmonise effect and degree of pitch shift (how many half steps to pitch shift). The harmoniser is also a good effect to manipulate, since it is easy to hear a difference in.

The vocoder effect's parameters would be the amount of effect, but since it needs an instrument and a voice input it is not an effect to consider.

The auto-tune effect's parameter is the amount of correction and which tones it corrects to in real-time. This effect could be also considered useful because it will be possible to hear a difference in it.

2.2 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.

2.2.1 TC-Helicon's Perform V

The TC-Helicon's Perform V is a vocal multi-effect processor that can be attached to a microphone stand as seen in figure 2.1[8]. It has three effect buttons, three preset buttons, a big knob and other buttons. The effects are reverb, echo, "double" which is a harmoniser, equaliser, compressor and several others. It is possible to download an application that can connect with the Perform V. The application offers many pre-made sounds and it has a wireless connection.

Figure 2.1: TC-Helicon's Perform V[8]

The Perform V is good for live performing if the singer has the processor in front of them on the microphone stand. The preset buttons make it easy to change effect quickly.

If the singer plays an instrument, it will be difficult to change effects without interrupting the playing of the instrument. Another downside is that the singer is limited to only three presets and only one knob to turn.

2.2.2 Electro Harmonix Voice Box

The Electro Harmonix Voice Box is a more advanced processor than the TC-Helicon[9]. 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 include different kinds of harmonies, unison-whistle and a vocoder, which the TC-Helicon does not have.

Figure 2.2: Electro Harmonix Voice Box[9]

The Voice Box has to be on a flat surface 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 a variety of effects and knobs - this can make changing effects and effect parameters difficult, even more if the pedal is on the floor.

2.2.3 Mi.Mu Gloves

The Mi.Mu Gloves are gloves made for making music and controlling sound[10]. They were created 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 the software of the gloves is open source which means it is free to use. 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.

Figure 2.3: Mi.Mu Gloves[10]

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, everyone can customise them to their own preferences and needs which results in both simple versions and complex.

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

In this study Yonezawa et al. made a glove that controls voice effects[11]. The wearer of the glove controls a puppet and makes hand gestures as seen in figure 2.4.

Figure 2.4: HandySinger Glove[11]

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, as seen in figure 2.5.

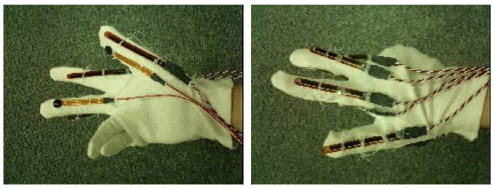


Figure 2.5: HandySinger Glove sensors[11]

The glove measures both forward bend and backwards bend. The gestures that the glove recognises 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.

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

The study by Maes et al. utilises body gestures to enhance a singer's voice[12]. The system is a human-computer interface that uses 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 can be seen in use in the figure 2.6. 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.

Figure 2.6: The One-Person Choir in use[12]

2.2.6 Discussion

The research of the state of the art has shown that there exist different kinds of products that all are able to manipulate a performer's voice. Many different pedals that can change effects exist, like the Electro Harmonix Voice Box, but such pedals often have to be on the ground out of reach[9]. The Voice Box has many effects and knobs, but this means it might be difficult to change effects and effect parameters quickly. On the other hand there is the TC-Helicon's Perform V, which enables voice manipulation from an easier to reach point - the microphone stand[8]. The Perform V has few, but big buttons and a big knob to turn, and this means the singer can change effects faster than with the Voice Box. Both units require the performer to look and spend attention on the interface in order to apply effects.

The Mi.Mu gloves allow a performer to manipulate a voice and create different sounds while moving around freely[10]. Studies by Yonezawa et al. and Maes et al. show that gesture control is a big part of singing[11][12]. Combining gesture control with the concept of the Mi.Mu glove could be interesting.

2.3 Gestures

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

There are three stages of a gesture: registration, continuation and termination, as seen in figure 2.7[13, pp. 127-134].

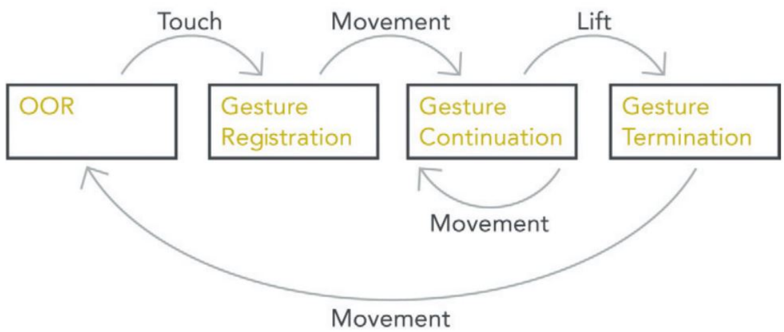


Figure 2.7: Out of Range(OOR) and the three stages of gestural input [13, pp. 127]

The registration stage is when the system registers that the user wants 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 if they want the gesture to be able to change the parameters or not. An example is 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 the possibility to differentiate between a lot of gestures is removed.

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

While the previously mentioned 'zoom' gesture is a rather good example of a gesture used to change parameters, a bad example of such 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 is that it implements no way to change the parameters dynamically. Another bad thing about it is that one has to specify a border between the action of simply moving something across the screen and 'flicking' something on the screen.

2.3.1 Discussion

The desire of this project is to not only apply effects, but also to be able to change the parameters dynamically during a performance. With this in mind, it is a requirement to include all three stages to make a fitting gesture for the different effects, since including the continuation stage is the only way to make this possible.

2.4 Conclusion

There are numerous sound effects each one unique in the way it works and is applied, so understanding their limitations or applications in a real-time setting is important.

By going through each of them and learning about them, the effects most relevant to the purpose of this project were 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. Most of their designs vary between a glove and a plug in device. Those design differences also provide different accessibility while performing.

Furthermore, 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 the gesture.

Finally, 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 use to apply an effect or change its parameters, is by pushing a button or turning a knob, as can be seen in figures 2.1 and 2.2.

2.5 Problem Statement

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

How can one design and implement a wearable device with an intuitive interface that applies voice effects in real-time using gestures, without the restraints of existing solutions?

2.5.1 Success Criteria

Based on the problem statement, some success criteria were specified which are as follows:

- The system has at least two effects
- Users use the correct gestures to change the effects
- The system performs the correct effect corresponding to the gesture, meaning it does not misinterpret

The system should have at least two effects to make it possible to test different effect parameters. It is important that the users perform the gesture that corresponds to the effect, since the evaluation is based on these effects working. It is also important that the system performs the right effect, whenever a gesture is performed. As an example, whenever a user performs the knob turning gesture, the system has to perform the effect that matches the gesture.

2.5.2 Minimum Implementation Requirements

The following describes the minimum implementations, which are needed for the prototype to be testable. These are specified by the authors of this project. The minimum requirements are as follows:

- The system must implement the use of an Arduino
- The system must implement at least one sensor 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 use audio from a microphone

3. Design

This chapter describes the design process for this project. Initially, the concept of the system is presented with the storyboard to explain the idea of the project. The system is then shown and described with the sketches and the iterations that it went through. Furthermore, the mental model lo-fi test along with an affordance scheme is explained, and the results of the test are shown. In the end of the chapter, the design choices for the final design are discussed and the chapter is concluded.

3.1 Concept

The concept of the product is to be able to apply voice effects and change their parameters in real-time without using a stationary panel or have someone backstage do it for you. The idea is based on giving the singer/performer more freedom and control over how they want to sound by using natural gestures to change the parameters. Most performers have one or both hands available during performances, so the concept is based on a controller made for hand movement. The controller would implement a gyroscope to sense the hand and wrist movement, the functionality of the gyroscope is explained further in the implementation chapter. The chosen gesture is based on which fingers the user connects. The idea behind the use of two fingers is to create a connection that initiates a specific effect. The index- or middle finger each control their own effect when the finger is connected to the thumb. The two original ideas for voice effects were pitch shift and reverb. Reverb was later changed to harmonisation beacuse it is easier to hear a parameter change in.

- Harmonising: This will be controlled by turning the hand while having the thumb and the index finger pressed together, like turning a knob or volume control.
- Pitch: This will be controlled by lifting or lowering the hand while having the thumb and the middle 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 device, see figure 3.1. This storyboard was part of the proof of concept that together with the initial sketches would explain the idea better.

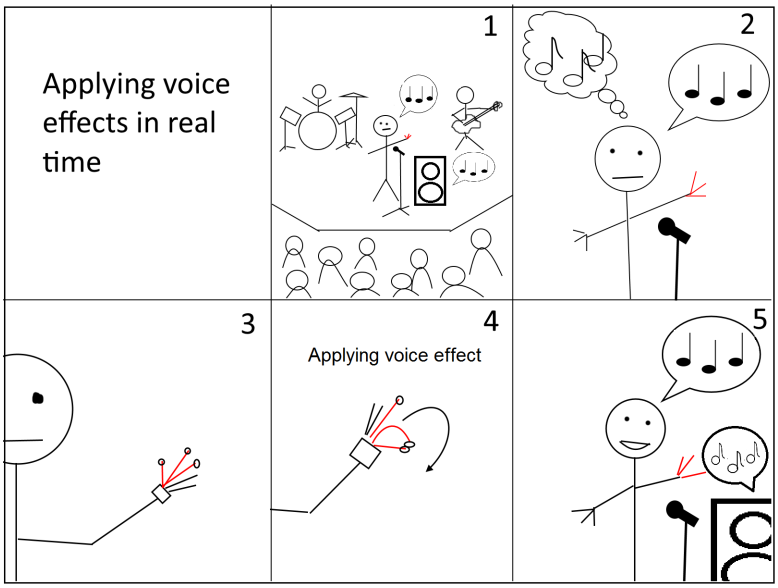


Figure 3.1: The Storyboard

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 has copper band on the 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.

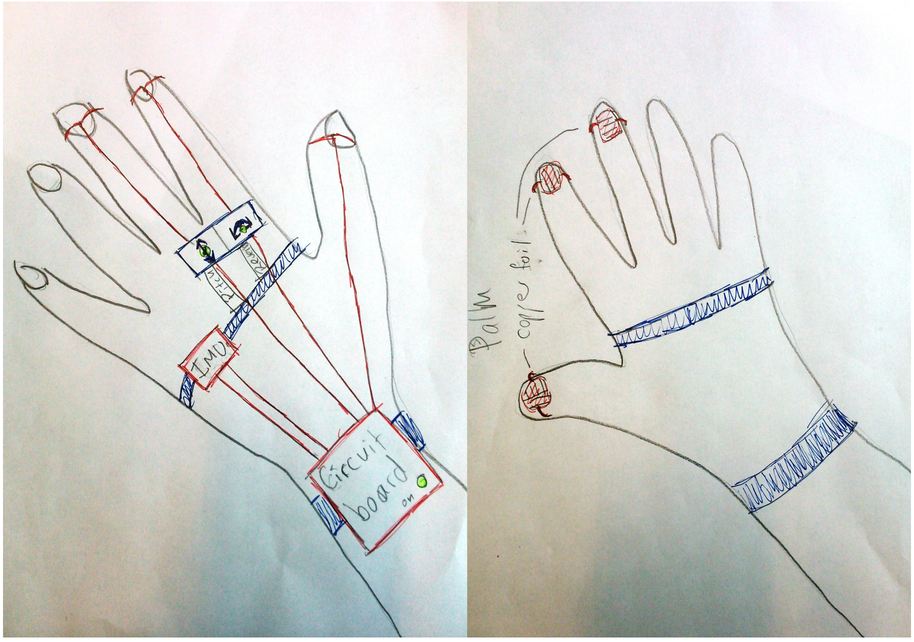


Figure 3.2: The initial sketches showing the idea of the system. The text on the right figure says: "Palm" and "Copper foil".

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 and the circuit board is attached to the hand by a velcro strip.

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 activation stage. None of them connected their fingers.
- Most eventually figured out which type of gesture had to be done, but only after some trial and error
- They all found out which finger created which effect

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.

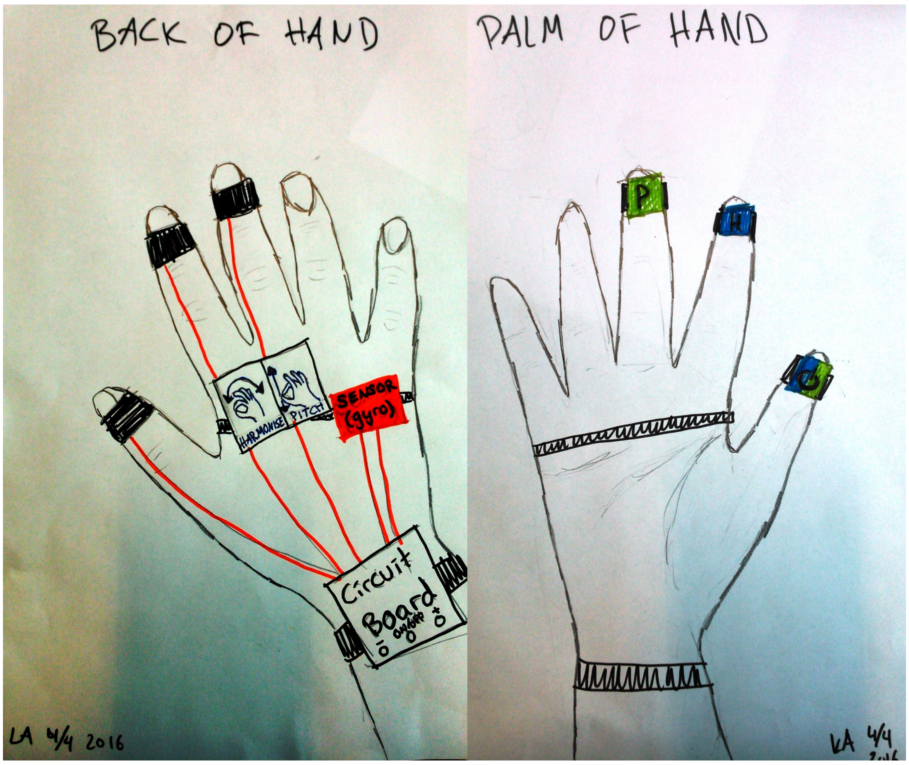


Figure 3.3: Second set of sketches after some changes

Colour was also added to the copper foil a different one on the index and middle finger and then both on the thumb.

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 change in the effect, with a minus and a plus sign.

A quick informal test was done with two participants with the revised sketch. This time 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.

- 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.

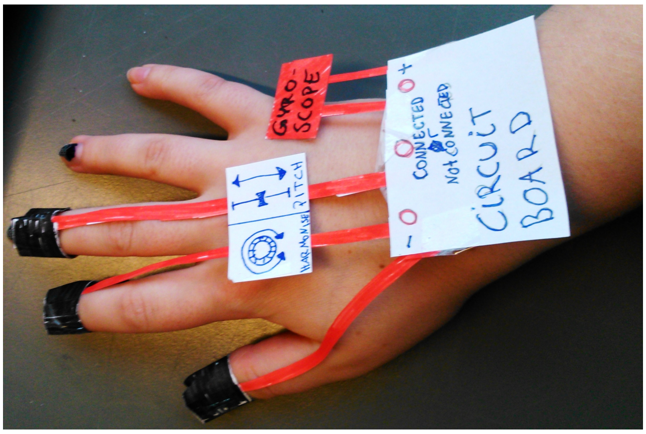


Figure 3.4: The Lo-fi of the system

3.4.1 Affordance Scheme

The affordance scheme is separated into three categories: perceived affordance, feed-forward, and feedback[14]. 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, one expects 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 system.

State	Perceived Affordance	Feedforward	Feedback
Inactive	It is a wearable glove and the copper plates can be connected	Harmonise and pitch can be manipulated with index and middle finger	All LEDs are OFF
Active		Perform gesture on labels = effect change. Release fingers = inactive	"Connected" LED is ON
Performing Gesture		Release fingers = inactive + no effect	Voice Changes and the plus and minus LEDs light accordingly

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 ilustrations 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 red writing in the table. Additionally, the parts which were generally understood by the users has the text coloured green.

State	Perceived Affordance	Feedforward	Feedback
Inactive	It is a wearable glove and copper plates can be connected	Harmonise and pitch can be manipulated with index and middle finger	All LEDs are OFF
Active		Perform gesture on labels = effect change and disconnect fingers = inactive	"Connected" LED is ON
Perform-ing Gesture		Disconnect fingers = inactive + no effect	Voice Changes and the plus and minus LEDs light accordingly

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. This could also be a result of bad simulation of the LED state changes in the test. Another thing that the users had a hard time understanding was the colour coding on the fingers. Most did not connect their fingers to initiate the effect change.

The things that the users perceived correctly was that the system was to be used as a glove and that each of the fingers corresponded to their own effect. After the alterations based off of earlier tests, the new illustrations had much more success regarding the gestures. Most understood the feedforward provided by the labels and performed the gesture correctly, although without connecting the fingers.

3.5 Design Choices

Throughout the design process a lot of choices have been made regarding the design of the system. The reasoning behind these decisions will be discussed in this section.

Very early on in the process, it was decided to make the gestures based of some gestures that people would normally associate with manipulating music, like a volume knob or a slider on an equalizer, as seen in figure 3.5.

Figure 3.5: The AX 301 Synthesizer Amplifier with knobs and sliders[15]

According to Wigdor and Wixon it is important to mimick real world physical interactions to create more natural experiences for the user[13, pp. 47]. The gesture of turning a knob was assigned to the harmonise effect, while the vertical slider gesture was assigned to the pitch shift. Since pitch shift is manipulated by making a higher or lower pitch, it makes sense to make the gestures simulate a vertical slider, with a higher pitch being assigned to moving up, and lower pitch assigned to moving down. Furthermore, in accordance to the "turning a knob" and "moving a slider" sentiment, it had to be implemented on the actual glove. This was done by adding some copper band on some velcro for the fingers. This copper was then connected to the circuit board with wires. When connecting the fingers, this would then allow the corresponding effect to be activated, and then the effect could be changed by using the gestures.

When looking at the feedback for the glove, other than the voice changing, it was decided to use LEDs as a way to communicate state changes. A red LED was used to indicate that the effects were being changed in the negative direction, e.g. pitch being lowered, and a green LED was used for the positive direction. Additionally, these LEDs would be accompanied by labels. A minus sign for the negative direction and a plus sign for the positive. The red LED and the minus sign would be placed on the left side of the circuit board to indicate the negative direction the "knob" had to move. The green LED and the plus sign would be placed on the right side to indicate the positive direction. Finally, a yellow LED was placed in the center of the circuit board. This LED will indicate whether the glove is active or not. Active being the state where an effect can be changed.

3.6 Conclusion

The first concept of the system was a glove that uses a gyroscope to sense hand movement, and apply voice filters according to the movement. A storyboard was created to show how the system would be used and in what context. The first sketch showcased the system with copper foils on the fingers, and labels to explain how to operate the device. The second iteration also included new labels to provide better feedforward and LEDs for feedback when interacting with the system. The third iteration once again included new labels, this time with more success and was made into a lo-fi model. This iteration was compared to the affordance scheme with the results showing the labels were better, but the system was lacking elsewhere, namely an indication of the connection of the fingers to apply an effect. The test participants also had some trouble understanding the feedback of the system, as in how the LEDs worked and what sort of feedback they actually provided.

4. Theory

This section explains the theory behind the effects that have been used in the implementation of the prototype.

4.1 Sampling

Sampling and quantizing is used as a way to convert analog signals into digital. Sampling is discretising the time variable. Meaning splitting the continuous signal into discrete signals, by isolating values at regularly spaced time intervals. Quantizing is the same operation but done for the pressure signal, which represents the displacement of a microphone's diagram from its resting place, these value ranges from -1 to 1.

When the signals have been converted into discrete signals they can be split up into samples, sample being a value or a set of values in time. The average amount of samples that each second is split up into depends on the goal and the use of the signal. This is called the sampling rate/frequency which is measured in Hz, and generally speaking the higher the sampling rate the wider range of sounds can be heard and the signal is of higher quality. However after it passes a certain threshold human ears cannot tell the difference.

If a sampling rate which is high compared to the frequency of the signal is used, the outcome signal is of good quality since there are several samples to represent each period of the signal. If an increasingly smaller sampling rate is used on the same signal then at one point there will not be enough samples representing each period to give the correct representation of the signal's frequency [16]. An example can be seen in figure 4.1 showing how a smaller sampling rate than the signal's frequency is used and creates a signal with lower frequency than the original.

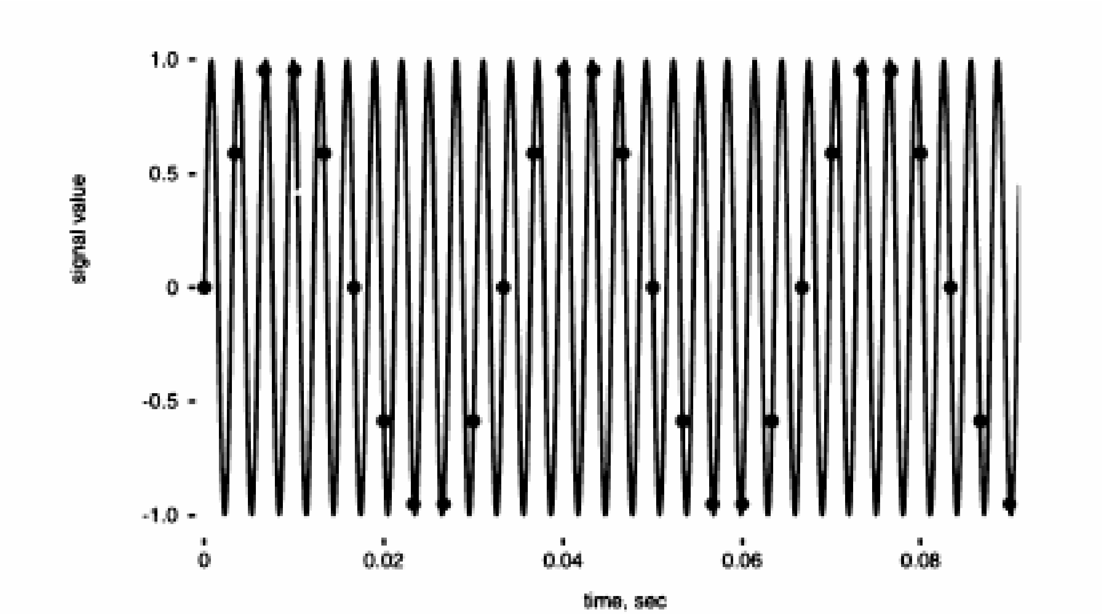


Figure 4.1: Sampling a 330 Hz signal at the rate of 300 Hz [16]

The most common sampling rates used when it is needed to capture the entire range of human hearing(20-20000 Hz) are 44.1 kHz(same as used in compact discs), 48 kHz, 88 kHz and 96 kHz. The reason for a sample rate to be twice as high as the signal's frequency is due to the Nyquist theorem which claims that a double rate is enough to perfectly reconstruct the signal. While it is possible to reconstruct a signal without using a double sampling rate, other constraints of the signal need to be known. For the prototype a sampling rate of 48 kHz was chosen. A higher sampling rate is not necessary as a higher sampled signal than 50 kHz does not provide any additional information to a human listener while a lower sampling frequency e.g. 44.1 kHz could be used with the same effects, but any lower than that should not be used.

4.2 Pitch Shift

The first pitch shifters were built based on rotating tape heads [4]. They were used in the 1960s to change e.g. radio interviews, where they could use it to speed up the interview, but still maintain the same pitch. Six tape heads were used to read the normal tape that needed to pitch shift at a speed independent from the tape speed. The speed of the tape head determined the duration of the sound, and the direction of the tape heads controlled the pitch shift. When the head rotated in the same direction as the tape, the pitch was lowered and if it rotated in the other direction, the pitch was increased, as seen in figure 4.2.

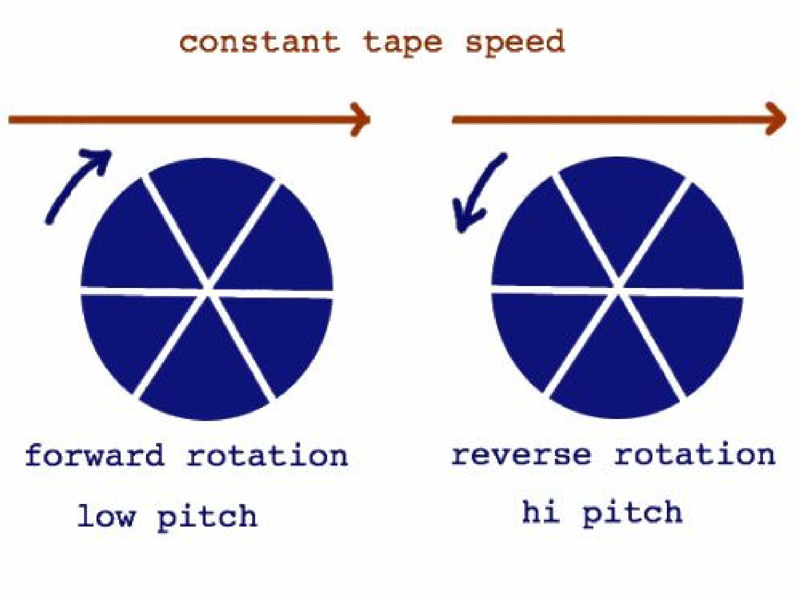


Figure 4.2: Graphical Impression of the tape head principle[4]

The first pitch shifters used six tape heads, but digitally you only need two. If you want to raise the pitch by e.g. three half tones, how much faster must the tape head go?

This equation converts the half tone number to a "tape head speed" number:

$$2^{h/12} = e^{0.05776*h}$$

where h is the half tone number, e.g. three half tones would give the "tape head speed" the number 1.189.

In a digital setting the tape heads are replaced with two read pointers that change their delay depending on the required pitch shift. This will be explained in the implementation chapter.

4.3 Harmonise

A harmony is a sound created by playing or singing different notes at the same time[17].

Singing harmony means that a backup vocalist or someone else is singing the needed notes in conjunction with the lead singer singing the main melody notes. In a digital setting this effect can be achieved with a single voice. This is done by pitch shifting the voice multiple times, each time with a different shift in half tones and stacking them together creating the effect of singing in harmony.

A chord or a harmony is multiple notes, usually three or more, played at the same time. A chord in which three notes are played is called a "triad"[18]. A triad is built on thirds, a root, a third, and a fifth note, see figure 4.3. The third being a third above the root and fifth being a third above third, or a fifth above the root. Depending on the quality of the two third intervals the quality of the triad changes.

Figure 4.3: A major, minor, diminished and an augmented triad [19]

In western music there are seven kinds of triads, three major, three minor and one diminished. The major triad and the minor triad will be implemented in the program of this project.

5. Implementation

This chapter describes the theory behind the technical aspects of the prototype and how they were implemented into the prototype. This includes the necessary hardware to make the prototype and how these were used in conjunction with the software to make the prototype work. The software includes both the operations of the hardware and the musical processing of the input sounds.

5.1 Making the Prototype

This section describes the hardware for the system and how it was made into a working prototype. This includes both the schematic, which explains how the system is put together, and the software for the prototype which both covers the software for the Arduino and sensor, and how they work together with Pure Data to process the input sounds for both harmonizing and pitch shift.

5.1.1 Hardware

The prototype consists of an Arduino Mega 2560[20], 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[21]. It has a tri-axis gyroscope, magnetometer, and accelerometer.

In this project, the gyroscope is the only one that is being used. It takes data from the sensor, that is then calculated by the Arduino program and grants the rotation of the hand in both the x- y- and z-axes.

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 are located on the thumb, index and middle finger. These are velcro based, with copper tape on them, to create the connection when pressed together.

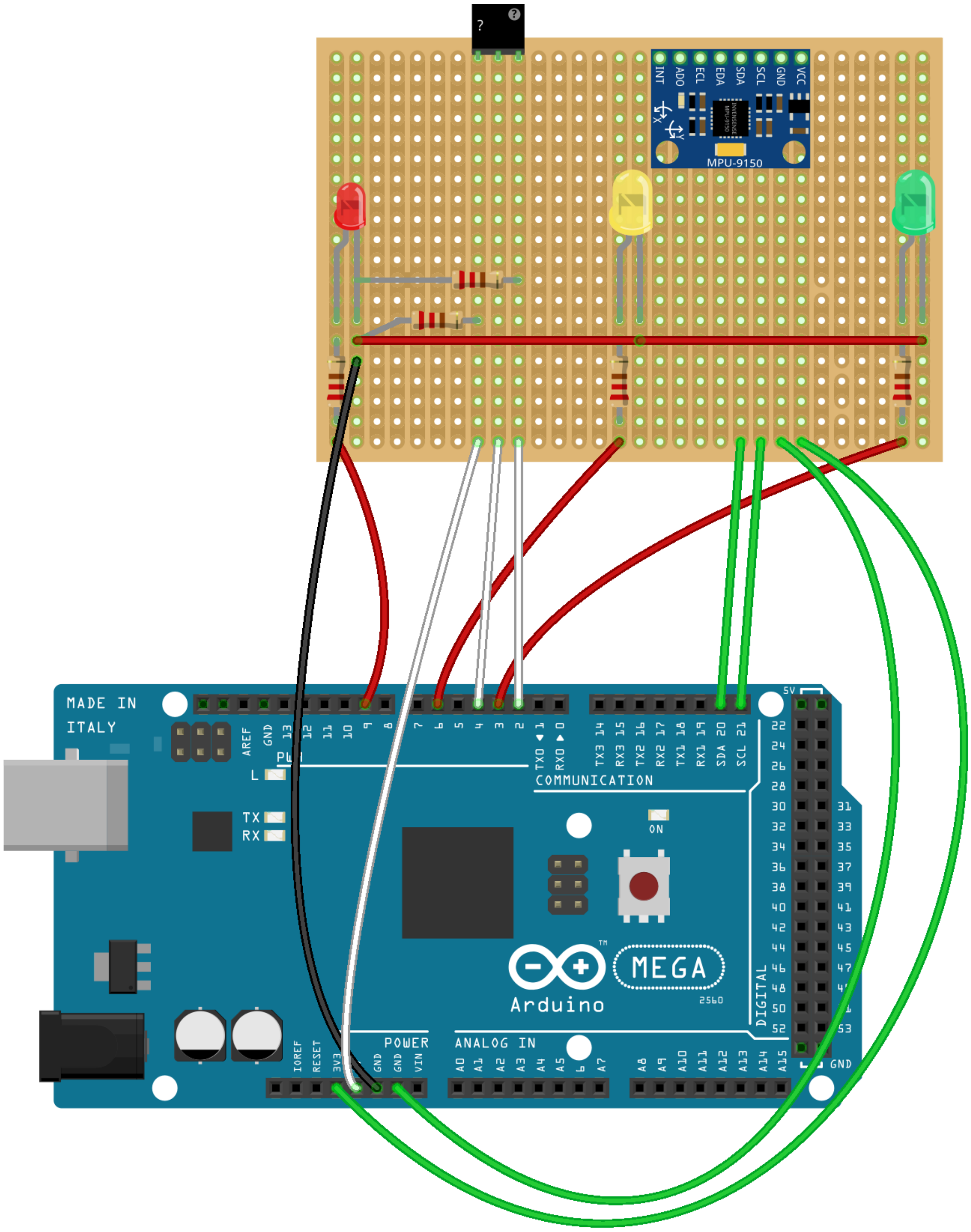


Figure 5.1: Illustration of the prototype's circuit board

The circuit board consists of three LEDs - a red, a yellow and a green. They light up under different circumstances, which are explained in the next section of this chapter.

The LEDs are connected to resistors of 220 Ohm, which are wired to the Arduino as seen in figure 5.1. The green LED is connected to pin 3 on the Arduino. The yellow LED is connected to pin 6 and the red is connected to pin 9. This is essential to make it work in the Arduino code, which will be described later.

The red wire connects all the LEDs and the black wire runs to the ground(GND) on the Arduino. The little blue circuit board on the yellow board represents the IMU. The necessary connections are: SDA, SCL, GND and VCC.

SDA is the data line that sends data between the two devices, and the SCL is the clock line that sends pulses at a regular interval[22]. Every time the SCL changes from low to high, a single bit of information is sent over the SDA.

The SDA 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 pin 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 pin on the Arduino. The white wire beside it, is connected to the index finger, and pin 4 on the Arduino. The last white wire is connected to the middle finger and pin 2 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 have to be connected to a pull-down resistor[23], which ensures 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. Figure 5.2 shows a schematic with a pull-down resistor.

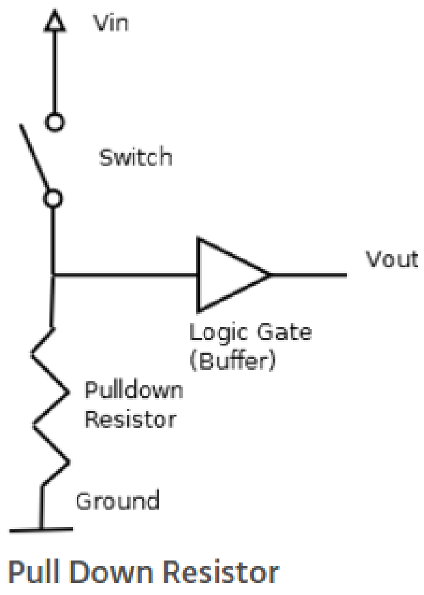


Figure 5.2: Schematic of pull-down resistor[23]

Figure 5.3 shows the prototype. As one can see, the velcro for the fingers has copper tape on it, which has been glued onto the velcro. The velcro enables the user to wear the glove and to use it properly.

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.

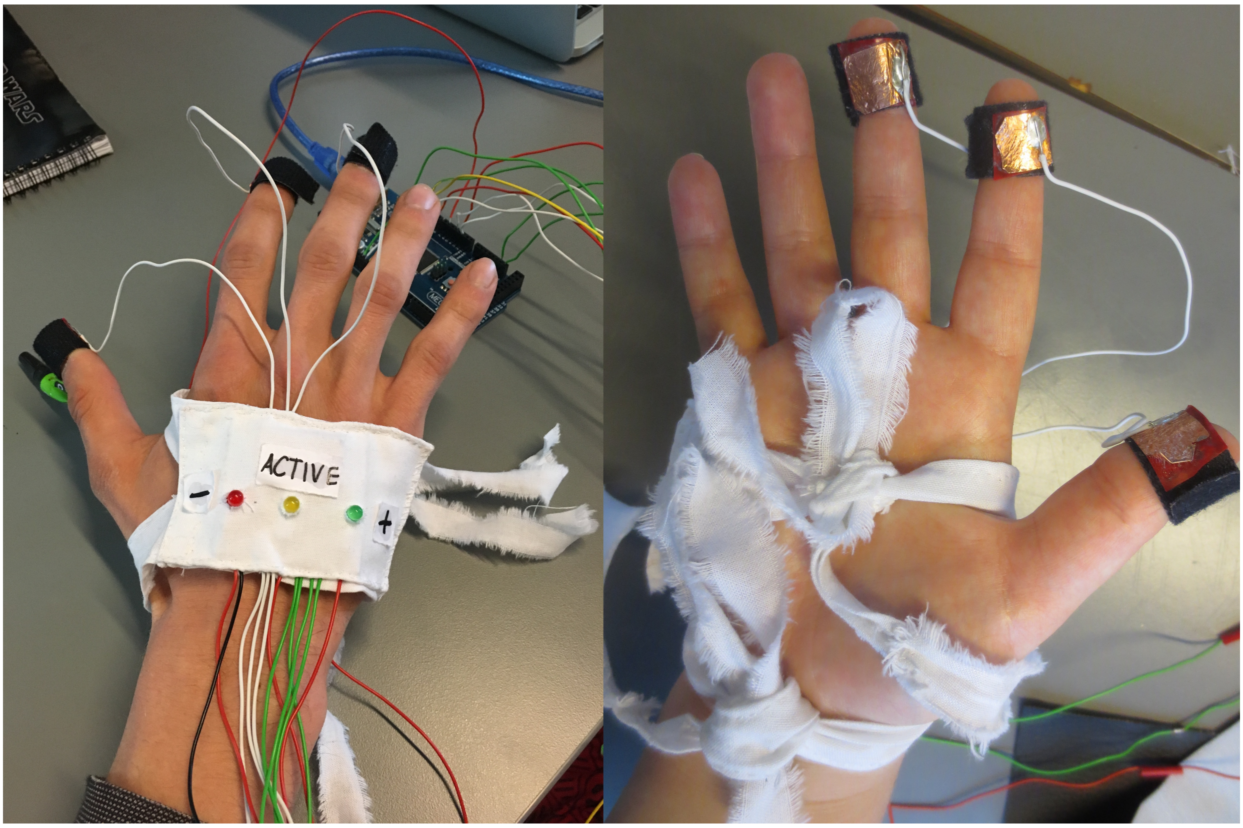


Figure 5.3: The front and back of the prototype

5.1.2 Software

This 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 [24].

Information from the IMU is continually read and sent to the Arduino program. This data is then calculated by the Arduino program. It provides the pose, the angle, and the current movement of the IMU, all based on the data from the gyroscope, accelerometer and magnetometer.

The first part of the code shown in figure 5.4 reads the state of the copper plates. The following if-statement is executed whenever the index finger is connected with the thumb.

If the user turns the hand to the left, the code sends the information to Pure Data, and if the hand is turned to the right, it sends that information.

The direction of the hand determines the pitch within a value of 0-255. The motion is a rolling motion, which resembles the motion of turning a knob.

```
//Reading copper plates/buttons
isOn2 = digitalRead(2);
isOn4 = digitalRead(4);


//Index Finger
if (isOn2 == HIGH) {

    if(toggle==false)
    {
     Serial.write(2);
     toggle=true;
    }

  if (pitch > 255) {
    digitalWrite(9, HIGH);
    Serial.write(255);
  } else if (pitch < -255) {
    digitalWrite(3, HIGH);
    Serial.write(-255);

  } else if (pitch < 255 && pitch > 0) {
    analogWrite(9, abs(pitch) * 3);
    digitalWrite(3, LOW);
    Serial.write(map(ardVal,0,255,10,255));
  } else if (pitch > -255 && pitch < 0) {
    analogWrite(3, abs(pitch));
    digitalWrite(9, LOW);
    Serial.write(map(ardVal,0,-255,-10,-255));
  }
```

Figure 5.4: Code describing what happens when the index finger is connected

The code used for the middle finger is almost identical. It does however use different pins. Initially it was the plan to use the raise/lower movement for the pitch shifting, but since it caused some errors, it was decided to use the rolling motion for pitch also.

Figure 5.5 shows the last part of the code. This is an else statement that makes sure that the button is off whenever the copper plate is released.

```
} else {
  digitalWrite(9, LOW);
  digitalWrite(6, LOW);
  digitalWrite(3, LOW);

  //Copper button is "pressed" once again.
  //This makes sure the button is off when the copper plate is released
  if(toggle==true){
   Serial.write(2);
   toggle=false;
  }else if(toggle2 == true){
    Serial.write(4);
    toggle2=false;
  }
  Serial.write(0);
```

Figure 5.5: Else statement which makes sure the button is off when released

Pure Data

The audio processing has been done in Pure Data(PD)[25], 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 make it possible to create different audio effects. This subsection describes how PD has been used in this project, and explains the patches made for this project.

The first figure shows the main patch of the PD program, see figure 5.6. The figure shows three coloured rectangles, red, green, and blue.

The content of the red rectangle is Arduino related. PD gets data from the Arduino from the "comport" object, which is a serial port interface.

The comport object needs a device number and a baud rate. In this case the device number is 1, and the baud rate is 4800 since this must be the same as the Arduino program's baud rate. The comport object takes three message objects as input: the name of the USB-device which starts the comport object, a message which shows which ports are available and a close message.

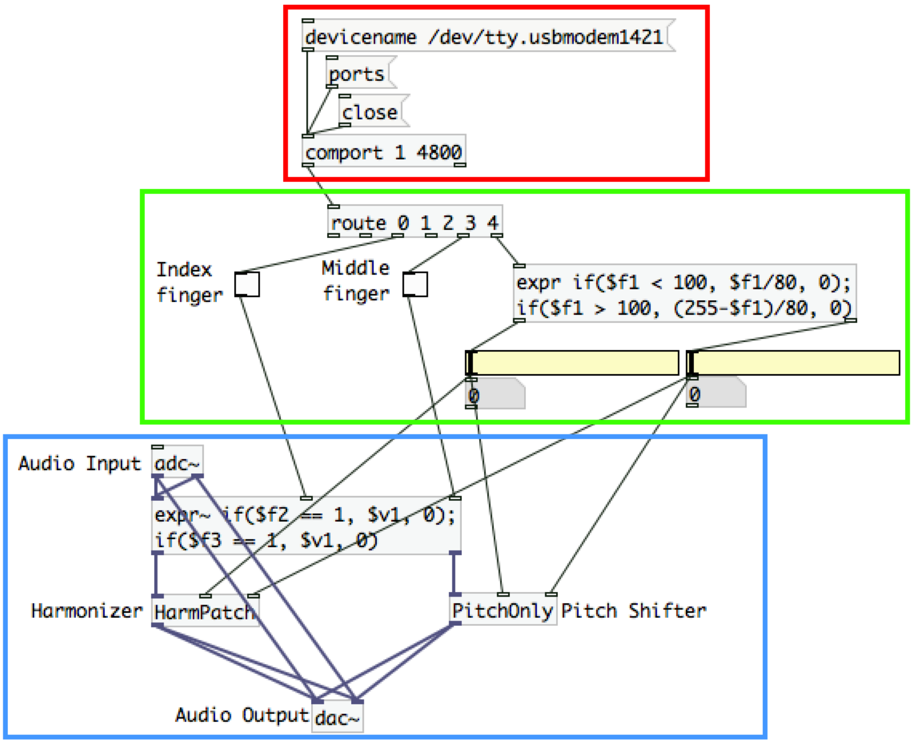


Figure 5.6: The main PD patch. Red rectangle is the Arduino part, green is the variable part, and blue is the effects part.

The green rectangle processes the output from the comport object. The "route" object takes the data from the comport and splits it up into five parts.

Route output number two and four correspond to the copper-plate buttons, e.g. if the index finger is connected, route output number two activates the toggle.

The last route output is the sensor data, which goes to an "expr" object. The positive sensor data is between 10 and 100, and the negative between 255 and 155.

The "expr" object splits the positive and negative values from each other, and divides them with 80. The negative values are subtracted from 255 first. The values go to two sliders.

The blue rectangle does the audio processing. Firstly, the "adc~" object which converts analog sound to digital takes the microphone input and sends into a "expr~" object. If the index finger is connected, the audio is sent to the "HarmPatch" patch, and if the middle finger is connected, the audio is sent to the "PitchOnly" patch. Both the "HarmPatch" and "PitchOnly" patches read the arguments of the sliders from the green rectangle. The outputs from the effects are then sent to the "dac~" object which converts digital sound to analog, which sends it to the sound output, in this case headphones.

Figure 5.7 illustrates the "HarmPatch" patch used for harmonising. As one can see in the figure, the audio coming from the "inlet~" object is sent to four "PitchShifter" objects. The "PitchShifter" object takes a message as an argument, which corresponds to the number of half tones one would like to pitch-shift. The two first "Pitchshifter" objects create a minor chord by pitch shifting by three and seven half tones. The last two "PitchShifter" objects create a major chord by pitch shifting by four and seven half tones. The outputs from the four "PitchShifter" objects are then added in an "expr~" object, which uses the sensor input from the green rectangle from figure 5.6. The output from the "expr~" object then goes to the "outlet" object.

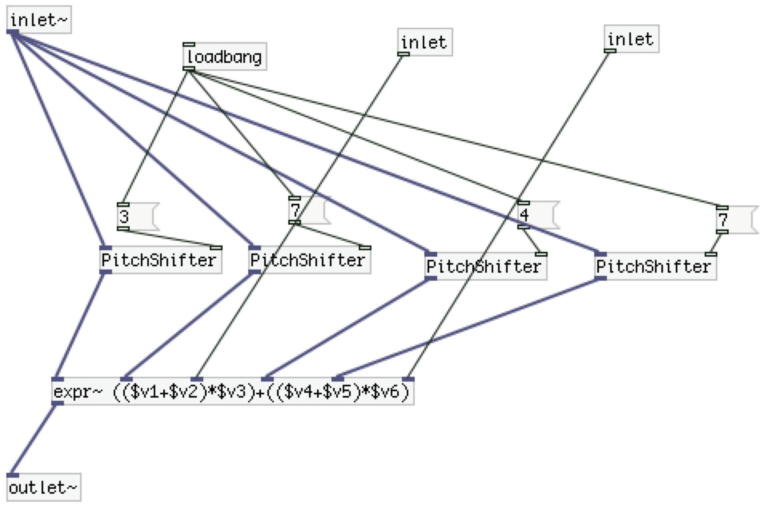


Figure 5.7: The "HarmPatch" patch. The audio goes to four PitchShift patches, and are added together again after

See figure 5.8 for the Pitchshifting patch used in the harmoniser effect. The green rectangle receives a message, the half tone number, but must first convert it to a value that PD will use later. The conversion uses this equation[26]

$$2^{h/12} = e^{0.05776*h}$$

where h is the half tone number. E.g. three half tones would give the converted number 1.189.

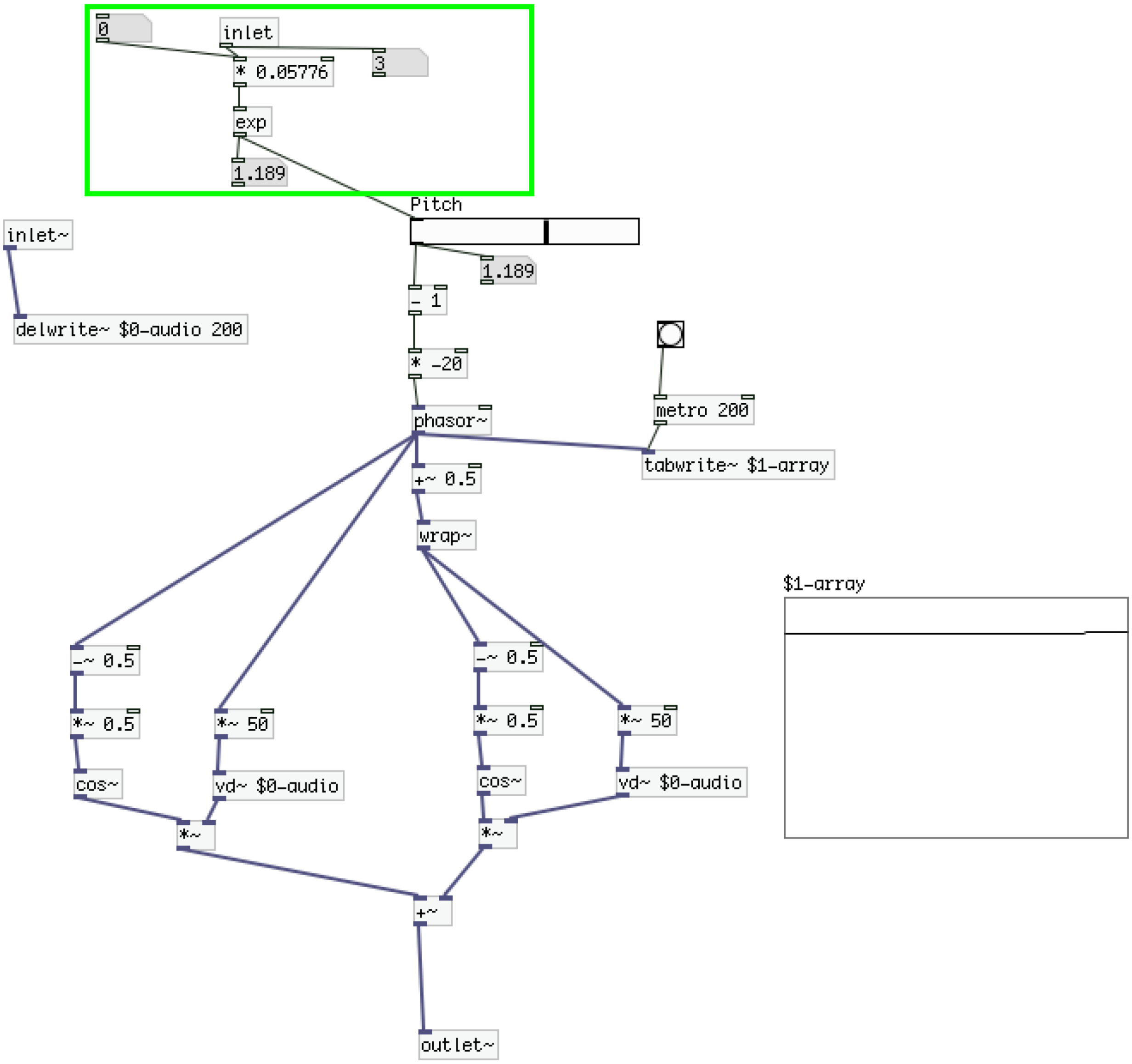


Figure 5.8: The Pitchshifting Patch used for the harmonize effect. The green rectangle is a bit different compared to the Pitchshifting effect.

The pitch shifting effect's patch does not convert numbers, but instead gets the slider data from the main patch, see the red box on figure 5.9. Since the pitch shifting effect takes the input from 0-2, where 1 is regular pitch, the slider data is either added or subtracted with 1.

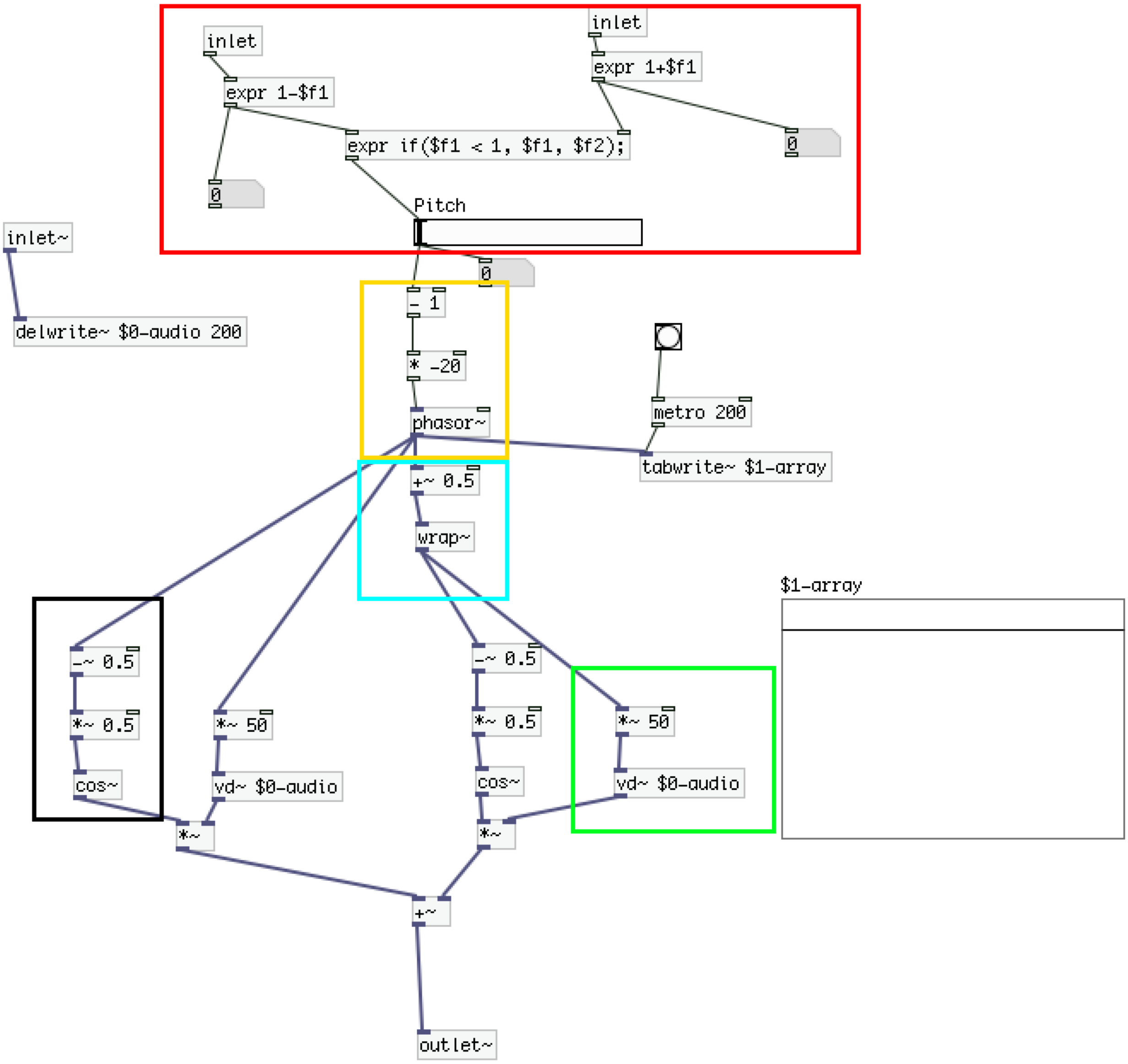


Figure 5.9: The pitch shifting patch

We find the delay by going through some steps. First, we find the difference between the "tape speed" and the original tape. Second, we multiply this number by -20. This is found by dividing the sample rate, 44100 Hz by the offset maximum, in this case 2205 samples, see the yellow square in figure 5.9. The "phasor" object, which makes a sawtooth, takes the number we have just found as an input, and sends the output to the two delays.

A phasor generates a sawtooth, and compared to the sinusoid, which goes from -1 to 1, the sawtooth goes from 0 to 1[27]. Normally, the sawtooth goes from low to high, and then abruptly to low again, as seen in figure 5.10. Since our tape head speed is going faster than the tape, it will progressively go to zero. In order to make the phasor fit with the delay, it must be inverted by making the input negative. It will go from high to low, and then to high again. This is why the 20 is negative in the yellow box.

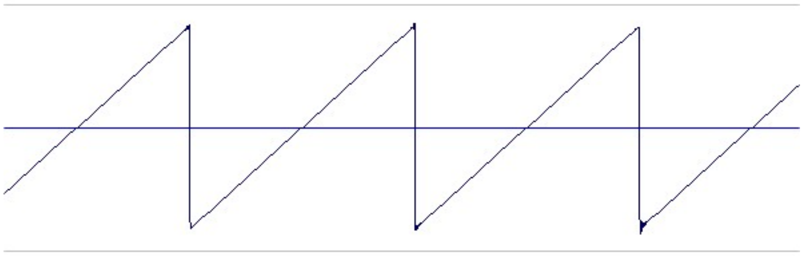


Figure 5.10: A Sawtooth wave[28]

The output of the phasor goes three ways. It goes to two variable delays, "vd~" objects, and two "cos" object. Before going to the "vd~" and "cos~" object on the right side of figure 5.9, one output is shifted half a cycle by adding 0.5. In order to shift it back between zero and one, a "wrap" object is used which forces it between one and zero as seen in the blue box in figure 5.9.

As mentioned before, the "vd~" objects are variable delays which get a varying delay time from the phasor. But before doing so the delay time is multiplied by 50, see the green box in figure5.9. This means that the "vd~" objects get a maximum delay time of 50 ms and a minimum of zero. A slight issue arises when the variable delay suddenly jumps because of the sawtooth wave and produces a click sound. In order to remove the clicking noise a smooth crossfade is used, which is done by using the "cos~" objects. Since the "vd~" and "cos~" objects on the right side of the figure5.9 are half a cycle ahead of the left side, when one variable delay is in the maximum or minimum delay, the other is in the middle. Furthermore, using the "cos~" function the extreme high or low delays are faded out and this is called cosine windowing, see the black box in the figure5.9. The "cos~" objects on both sides are then multiplied by its "vd~" objects" and the results are then added and sent to the outlet. This leaves us with a pitch shifted output that has no clicks.

6. Evaluation

The following chapter describes the evaluation and introduces the results from the testing of the prototype.

There were conducted two tests, an internal and a user test. This was done to test whether the prototype was functional on a technical level and to test whether it is usable.

6.1 Internal Test

To test the accuracy of the prototype, an internal test was carried out. The following hypotheses have been stated for the internal test:

H_0: The prototype works less than 90% of the time.

H_A: The prototype works more than or equal to 90% of the time.

The test consisted of the copper plates being connected 25 times each. First the index finger was tested with the thumb and afterwards the middle finger with the thumb. Whenever the yellow light on the system lit up it was considered a success. If the light did not turn on it was considered a failure. The times it succeeded and failed were counted.

The results ended up being as follows:

- Index finger connection test: 23 out of 25.
- Middle finger connection test: 25 out of 25.

The results have been analysed statistically using R[29]. R is a language used for statistical analysis. All the statistics have been calculated with a probability of success, p, of 0.75.

The results from the index finger connection test showed a p-value of 0.03211.

Whenever the p-value is less than the 0.05 significance level, a conclusion can be drawn from the test. In this case the H_0 is rejected for the index finger, since the test gave a probability of success of 0.92(92%). The prototype worked at least 90% of the time when connecting the index finger and thumb.

The results for the middle finger connection test showed a p-value of 0.0007525. The p-value lies below the 0.05 significance level, so a conclusion can be made. The probability of success for this was 1(100%). This means that this connection works at least 90% of the time, and the H_0 is rejected for the middle finger also.

Another internal test was carried out to test the accuracy of the effects. The test was conducted in a similar way, the effects were applied and then used. As long as the effect worked and changed the voice input, it was considered a success.

In the first test, the system worked 18 out of 25 times for both gestures. Since the gesture effects showed the same results, the p-value is the same for both gestures. The p-value ended up being 0.7265 which is not lower than the 0.05 significance level and therefore fail to reject H_0 meaning no conclusion can be made.

During the first tests, the system had a loose wire. The issue was fixed and the tests were then conducted once again with much better results:
All the tests worked 25 out of 25 times.

This gave a p-value equal to 0.0007525 for all tests, which is much less than the 0.05 significance level and a probability of success equal to 1(100%) and therefore H_0 is rejected for all the tests.

6.2 User Test

When conducting the user test, it was the intention to get some data from the users about the different effects and gestures.

6.2.1 Evaluation Plan

To evaluate the working prototype three focus points based on the device were stated. They are as follows:

1. Participants understand the link between gesture and effect
2. The design is awkward to wear
3. The participants will have no clear preference of effect

To evaluate the focus points a test was conducted with the goal to either prove or disprove them. First the participants were asked to sign a consent form which together with the script can be found here: Appendix A.1 and A.2. The test then consisted of the participants doing different tasks, which was followed by a few questions. The questions as well as the results of those questions can be found in Appendix A.3.

6.2.2 Apparatus and Setup

The following equipment was used to perform the test:

- Two laptops
- Zoom H4N microphone
- Headphones
- The Prototype

The following figure 6.1 shows the setup of the evaluation. One person wearing the system, a facilitator, and one person that makes sure the system works and holds the microphone (technician).

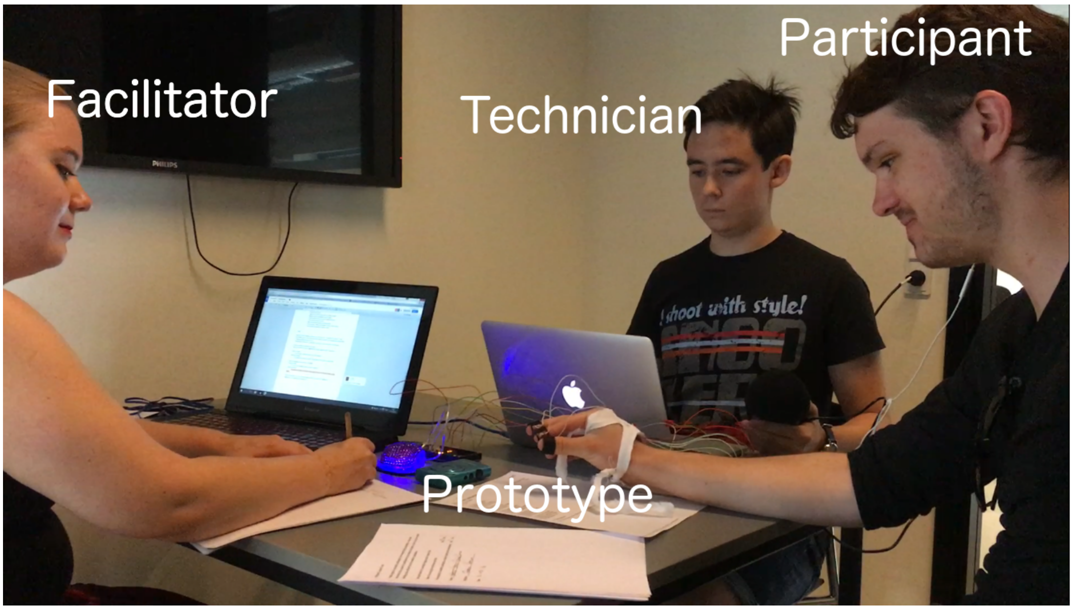


Figure 6.1: The Test Setup. There is a facilitator, a technician, and a participants. The technician has the microphone in his left hand.

6.2.3 Results

The first focus point deals with the link between the gesture and the effect. Here the question were asked: "Did the gesture make sense together with the effect?" All the participants answered yes to this and also stated why. A female participant said “[it is] like turning a button up and down”. A male participant mentioned that “It is like turning [the volume] up and down". Another male participant said "[it is] like turning a knob". These three are examples of the exact idea the gesture was based upon.

The second focus point deals with the physical aspects of the prototype. Here the question asked was: "Was the device comfortable to wear?" Again all participants agreed. The device was comfortable to wear.

The third focus point deals with preference of effect. The participants were asked: "Which effect did you prefer?" All participants, except one, answered pitch-shifting. This presents a clear preference for the pitch-shifting effect.

Aside from what the quantitative data from the questions, there were some additional data. Mainly the comments and feedback from the participants during the test which is analysed as qualitative data. Especially the participant with a musical background gave a lot of useful feedback regarding the prototype. Four out of the ten participants stated that they could not hear a difference between the major and minor harmonies. The participant with a musical background heard the difference clearly. However an assumption could then be made that the other participants could not hear the difference since they did not know what they were listening for.

They could all clearly hear the difference between shifting the pitch up or down. However five participants mentioned that the pitch-shifting is very sensitive and that there is not a lot of middle ground between no effect and full effect.

Another thing mentioned by three of the participants is that even though all participants found the gestures intuitive, holding them for longer periods of time could be very straining and cramp-inducing on their hand.

When asked whether they thought the prototype useful or not, nine of the ten participants said that they thought it could be. Some said that it could be useful in theatre performance and another said that it gives more control to the singer. Both of these are some of the ideas this project is based on.

The one participant who had a musical background also gave some additional constructive feedback. She stated that the harmonise effect lacks the ability to be able to choose the intervals you want. She said that it would be "incredibly smart if the first notes added the 3rd interval and then the 5th interval, and maybe even removed the 3rd interval if you twisted the wrist even further. Then you are in more control of the notes and the expression the music makes." Changing the intervals is something to consider for further iterations.

6.3 Discussion

Several things about these tests must be considered before any conclusions are drawn. Only one participant from this user test fits this project's target group. This affects the results quite a bit since participants of this test do not have the necessary knowledge to know how the prototype could be utilised, and therefore cannot give very useful feedback on the musical aspect of the prototype. Then there is the issue that the first half of the participants tested a prototype with a loose wire which made it act oddly and made it more prone to not react to the gestures made by the participant. The questions were also mostly yes/no questions which gives a rather limited view of the participants' reactions and views of the prototype. A Likert scale would maybe have been more appropriate for more nuanced answers.

6.4 Conclusion

Based on the results from the internal and user tests, it can be concluded that the technical parts of the prototype works as intended. Even though the system has few features, it works every time and all the participants like the concept and thought it useful, even though they may not have enough knowledge to deem it so. Using the feedback that was gathered, certain adjustments could be implemented, i.e. less sensitivity to the pitch-shifting and more adjust-ability to the harmoniser. The concept of the prototype working with the hands works very well although further exploration of different gestures and effects could be a good idea.

7. Discussion

The problems this project address is that singers or performers cannot apply voice effects easily. The problem analysis chapter covered which kind of voice effects exist today, current voice effect pedals and interfaces, studies regarding voice effects and gestures, and what a gesture is and how it can be used.

There are several kinds of effect interfaces, some are on the floor and highly customisable but hard to reach/change while performing. Some effect interfaces are small and can e.g. be on a microphone stand, but are hard to customise.

During state of the art research, it was found that there are studies focusing on using gestures to change effects, and a project that focuses on a highly advanced system. The studies focusing on gestures that change effects, "HandySinger" and "One Person Choir" found that having gestures which make sense to the user is both possible and beneficial. Additionally, a project with a very similar concept was discovered, the Mi.Mu glove. Their project was focused around making a glove for musical interaction with a wide variety of features. In order to differ from their project, we decided to focus our project around making a glove with less features, but much more focus on making it intuitive for the user.

Many voice effects exist. Instead of turning an effect on and off, changing an effect parameter could sound better or make more sense. We chose the effects "harmonise" and "pitch shift" because they could benefit from continuous change, and because a parameter change would be easier to hear compared to an effect like the reverb. This does not mean effects like reverb, delay, vocoder etc. are bad effects to change parameters to, but rather effects that could be added in further development.

When deciding on a fitting gesture for the different effects introduced during the problem analysis, it was important to think about each effect individually. Should they just turn on and off or are there parameters that needs to be changed? Which parameters needs to be changed and in what way? In the case of our project, it was important to be able to change the parameters of each effect. The pitch shift needs to be able to lower and heighten based on the user's wishes. The harmonise should be able to change between minor and major, and the degree of which it can be heard. It was decided the most fitting gesture for pitch shift was to simulate a vertical slider, moving up to highten and down to lower. The gesture for harmonise was to simulate a knob, like when changing the volume on an old radio.

The prototype was designed to be worn on the right hand on the thumb, index and middle fingers. At this point, the prototype must be connected to an Arduino placed next to a computer with a lot of wires, which means the user has limited movability. We also found that there was no room for effect icons even though we had drawn them in the Lo-Fi prototype. This will be considered in further development.

The prototype only understands the knob-turning gesture. We did manage to get the vertical slider gesture working, but it was not stable enough to include in the evaluation tests. At first, the sampling rate of the Pure Data program was at 44.100 Hz, but it was found that it created unwanted audio problems sometimes, such as a delay, and was fixed by using a sampling rate of 48.000 Hz.

7.1 Quality of Solution

This section will discuss the qualities of the system. Our first success criteria states that we have to implement at least two effects. The prototype can process audio, more specifically take the microphone input and apply two effects, whose parameters users can change. The two chosen effects ended up being a harmoniser and a pitch-shifter, because they are easy to hear, and would benefit from gesture manipulation. A reverb effect was also considered but later discarded, since we felt that the effect did not benefit as much from parameter change in a live performance.

The second success criteria stated that the user should use the correct gesture for the intended effect. This was based off of the fact that we wanted to implement different gestures at the time. During the implementation it was discovered that the slider gesture worked, but was not reliable. So we decided to use the knob-turning gesture for both effects. Given this, the third success criteria, stating that the system should not misinterpret between gestures, was not fully achieved. Beyond that, the system had no problem recognising the knob-turning gesture.

The system fulfils all of the minimum implementation requirements by using an Arduino, utilising a sensor that works with the Arduino, implementing the system using audio processing in PD and connecting it with the Arduino software, and using a microphone to record the audio.

Even though the success criteria are approximately fulfilled there were some issues with the prototype. The prototype has many long wires attached to the Arduino, which is attached to a PC, and can accidentally disconnect from the Arduino. This limited the movements of the participants of the test, which might have affected the results of the test.

A problem that was encountered was that some users using the prototype for a long period of time, started to feel discomfort in their palms. Additionally, the users missed a source of feedback when turning their hand over too much, causing the LEDs to disappear from view. This might have affected their understanding of what was going on. Another thing to consider is the plus and minus icons that were designed to indicate what was happening with the parameters when tilting the hand. Plus and minus is not necessarily the best signs for both of the effects, but it was decided to go with this since it was hard to find a uniform indication that fit both effects. This may not make sense to some of the users.

7.2 Validity and Reliability

A problem we acknowledge is the fact that nine out of ten users that tested our system were not singers or performers. They had basic knowledge about some effects, but no experience performing. This means we cannot consider their answers, about whether it would be beneficial for a performer, valid, but rather look at their answers from a usability standpoint. On a note, the nine out of ten participants were fellow Medialogy students, which means they may be a little biased regarding the topic of the project. We had one participant, a non-Medialogy student, who is within our target group and therefore can be considered valid. However one participant cannot be considered reliable, since reliability requires replication.

Regarding the actual prototype, the soldering on the circuit board and copper plates on the fingers was not optimal, and had to be fixed a couple of times. This resulted in some users trying a prototype that was faulty which also might have affected the results of the test.

The test was conducted in a small room and the user was using headphones to listen to the output of the system. The ideal setting would be on a stage and with a normal performance setup - this would be a dynamic microphone and speakers so as to avoid a feedback loop. Furthermore, the user was the only one hearing his or her voice with effects, meaning no one conducting the test could hear the output. By watching the PD program and LEDs it could be observed whether or not it was working properly.

7.3 Further Development

During evaluation several ideas arose on how to improve the prototype. Some of them were obvious, for example making the prototype wireless. This would make it easier to manage and manoeuvre. Another rather obvious improvement would be to add more effects on the remaining fingers. This could be done by using different kind of sensors, e.g. flex sensors on the fingers. Which effects these are supposed to be is not clear but the possibility is there. More minute control of the effects would also be an improvement. This would mean that there would be a more gradual change to changing the pitch or more control of which harmonies are added by the harmoniser. An improvement of the design would also be on the table. A more easily wearable design, that one person could put on alone, would be a lot more useful and comfortable. A design that uses fabric similar to the Mi.Mu glove introduced in the State of the Art could be considered. It would also be interesting to see if the colour of the LEDs would matter or not.

8. Conclusion

The goal of this project was to make a wearable device that would apply voice effects in real time, and have an interface that was intuitive. Since most current effect units are stationary, and difficult to customise while performing, we wanted to make a better alternative. Based on this, we made the initial problem statement: *How does one create a wearable device that applies voice effects to a voice in real-time, while also having an intuitive interface design?*

During the research, different effects were examined and it was discovered that some effects benefit more from parameter change than others, among these were pitch shift and harmonisation, which were chosen for this project. Furthermore, inspiration was drawn from the state of the art, both regarding the usefulness of existing solutions and their shortcomings. The state of the art research also introduced gesture control and how it could be combined with singing. Several studies introduced a wearable device for use on a hand. Further research about gestural input was done, and we learned how to create an optimal gesture for a specific action.

Following the research, the final problem statement was defined as follows: *How can one design and implement a wearable device with an intuitive interface that applies voice effects in real-time using gestures, without the restraints of existing solutions?*

Based on the existing solutions, we chose that device should be worn on the hand. Two gestures were chose, based on real world gestures to change effect parameters, the knob-turning gesture and a vertical slider gesture. During the design process, a lo-fi prototype was tested. The test was based on the mental model concept and the feedback from the test showed that the participants did not entirely understand the interface at that point. The design was improved based on the feedback.

The theory chapter introduces the theoretical background for each of the effects that are later implemented. In this project, we use the tape head principle to implement the pitch shifting effect. The harmoniser will add two pitched voices in addition to the original voice, so you will hear either a major triad, or a minor triad.

The prototype consists of a glove, that utilizes a gyroscope sensor and copper plates connected to three fingers, which acts as buttons. The glove is connected to an Arduino, and the Arduino transmits data to the audio processing program made in Pure Data. The vertical slider gesture was not working properly, so the only gesture that was implemented was the knob turning gesture.

The user evaluation of the prototype was not performed with actual performers with the one exception. This means we cannot conclude if the prototype is an improvement over the normal effect units. We can conclude that the participants knew how to use the prototype, liked the knob-turning gesture, and they could see it used in a performance setting. An internal test of the technical aspects of the prototype was conducted, and based on the results of those tests, it can be concluded that the actual prototype is fully functional.

Bibliography

[1] J. Tyrangiel, “Singer’s little helper.” http://web.a.ebscohost.com/ehost/detail/detail?sid=d06a0fe3-92e7-46d8-92b4-c3646f32137c%40sessionmgr4002&vid=0&hid=4112&bdata=JnNpdGU9ZWhvc3QtbGl February 2009.

[2] C. Loeffler, “A brief history of time-based effects (part 1).” http://www.harmonycentral.com/articles/exploring-time-based-effects, September 2014.

[3] N. Redmon, “A bit about reverb.” http://www.earlevel.com/main/1997/01/19/a-bit-about-reverb/, January 1997.

[4] Katjaas.nl, “Pitch shifting.” http://www.katjaas.nl/pitchshift/pitchshift.html.

[5] M. Dolson, “The phase vocoder: A tutorial,” *Computer Music Journal*, vol. 10, no. 4, pp. 14–27, 1986.

[6] A. Hadhazy, “What’s auto-tune and how does it work?.” http://www.livescience.com/11046-auto-tune-work.html, September 2010.

[7] Innovativesynthesis.com, “Introduction to vocoders.” http://www.innovativesynthesis.com/introduction-to-vocoders/.

[8] TC-Helicon.com, “Perform v.” http://www.tc-helicon.com/en/products/perform-v/.

[9] EHX.com, “Vocal harmony machine/vocoder.” http://www.ehx.com/products/voice-box.

[10] 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.

[11] 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.

[12] 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.

[13] D. Wigdor and D. Wixon, “Brave nui world.” http://www.gm.fh-koeln.de/%7ehk/lehre/sgmci/ss2015/Literatur/Wigdor_Wixon_-_Brave_NUI_World.pdf, 2011.

[14] V. Kaptelinin, “Affordances.” https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/affordances.

[15] AudioFanzine, “Ax 301 synthesizer amplifier.” http://en.audiofanzine.com/keyboard-amplifier/moog-music/ax-301-synthesizer-amplifier/medias/.

[16] K. Steiglitz, *A Digital Signal Processing Primer with Applications to Digital Audio and Computer Music*. Addison-Wesley, 1996.

[17] “Music theory: Harmony part one (basics).” https://www.mymusicmasterclass.com/blog/music-theory-and-composition-basics-pt-1/.

[18] MusicAwareness.com, “7 harmonies tutorial.” http://www.musicawareness.com/Tutorial3.html.

[19] Johnston, “Triads – major and minor.” http://blogs.yis.ac.jp/johnstonb/theory-2/triads-major-and-minor/.

[20] “Arduino mega 2560.” http://www.arduino.cc/en/Main/ArduinoBoardMega2560.

[21] Sparfun, “Sparkfun 9 degrees of freedom.” https://www.sparkfun.com/products/retired/11486.

[22] Arduino.cc, “Master writer/slave receiver.” https://www.arduino.cc/en/Tutorial/MasterWriter, August 2015.

[23] “Pull down resistor.” http://playground.arduino.cc/CommonTopics/PullUpDownResistor.

[24] Arduino.cc, “Frequently asked questions.” http://www.arduino.cc/en/main/faq.

[25] PureData.info, “Pure data.” https://puredata.info/.

[26] M. Puckette, “Pitch shifter.” http://msp.ucsd.edu/techniques/v0.11/book-html/node125.html, 2006.

[27] FlossManuals.net, “Oscillators.” http://en.flossmanuals.net/pure-data/ch018_oscillators/.

[28] “Sawtooth wave generator.” http://vhdlguru.blogspot.dk/2015/04/sawtooth-wave-generator-in-vhdl.html, April 2015.

[29] T. R. Foundation, “Introduction to r.” https://www.r-project.org/about.html.

A. Appendix

A.1 - Script

A.2 - Consent Form

A.3 - Results from Questionnaire

A.1 Script

Script:

Welcome. We are testing our device as you can see here. This device's purpose is to add voice effects to your voice while performing, instead of having it pre-programmed, using pedals or having someone else do it for you.
The two chosen effects are "Harmonise" and "Pitch" as you can see here on the device.

First we would like you to put on the device. (this is how to do it)
Then we would like for you to sing/talk while trying to apply the effect "Harmonise"

- Do you hear a difference after you did the gesture?

Now try changing the pitch while you sing/talk.

- Do you hear a difference after you did the gesture?

Now we would like you to fill out this questionnaire and if you have any feedback not addressed in the questionnaire please tell us.

A.2 Consent Form

Participant Consent Form

I hereby agree to voluntarily participate in the research being undertaken. I have been informed that the consent to participate can be revoked at anytime, or that I can have certain identifying data withheld from the research publications.

I give permission for recorded images and/or videos to be:

-Taken of me during the interview and trial: Yes ☐ No ☐

-Re-produced in scientific publications, student reports and presentations: Yes ☐ No ☐

Name: ________________________________

Signature: ________________________________

Date: / /

A.3 Results from Questionnaire

Gender	Age	Do you sing?	Have you ever performed with voice effects applied?
Female	18-25	Yes	No
Male	18-25	No	No
Male	18-25	No	No
Female	18-25	Yes	No
Male	18-25	No	No
Female	18-25	Yes	No
Female	18-25	No	No
Male	18-25	No	No
Male	18-25	No	No
Male	18-25	No	No

Was the device comfortable to wear?	Did the effects work as they were supposed to?
Yes	Yes
Yes	Yes
Yes	Yes
Yes	No
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes

Did you like the effects?	Which effect did you prefer?
Yes	Pitch-shifting
Yes	Pitch-shifting
Yes	Harmonise
Yes	Pitch-shifting
Yes	Pitch-shifting
Yes	Pitch-shifting
Yes	Pitch-shifting
Yes	Pitch-shifting
Yes	Pitch-shifting
Yes	Pitch-shifting

Did the gesture make sense together with the effect?	Did you like the concept of the device?
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes

Do you think it would be useful to performers?	Why?
Yes	If they can wear it on stage. Maybe in theatre.
Yes	
Yes	It could be cool
Yes	The harmonise would be useful. I see no use for the pitch for performers.
Yes	It is easy to apply effects.
Yes	It gives more control to the performer.
Yes	It could be useful in some cases such as theatre.
Yes	Maybe it should be moved from the hand to the microphone. Other solution needed than pressing the fingers together.
No	There already exists so much else in this category.
Yes	One just needs to adjust and then it would be useful.