

---

---

# **Titel**

---

---

Project Report  
Group MTA 16440

Aalborg University  
Media Technology  
Rendsburggade 14  
DK-9000 Aalborg





**Media Technology**  
Rendsburgsgade 14  
DK-9000 Aalborg

## AALBORG UNIVERSITY STUDENT REPORT

**Title:**

Hand Recognition Game Control using  
Image Processing

**Abstract:**

Abstract

**Theme:**

Sound Computing and Sensor Technol-  
ogy

**Project Period:**

Spring Semester 2016

**Project Group:**

MTA 16440

**Participant(s):**

Alex Bo Mikkelsen  
Allan Schjørring  
Daniel Agerholm Johansen  
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

<b>Preface</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
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
<b>2 Problem Analysis</b>	<b>3</b>
2.1 Research . . . . .	3
2.1.1 Effects . . . . .	3
2.1.2 State of the Art . . . . .	4
2.1.3 Gestures . . . . .	8
2.2 Problem Statement . . . . .	9
2.3 Minimum Implementation . . . . .	9
<b>3 Design</b>	<b>11</b>
<b>4 Implemented Theory</b>	<b>13</b>
<b>5 Implementation</b>	<b>15</b>
<b>6 Evaluation</b>	<b>17</b>
<b>7 Discussion</b>	<b>19</b>
<b>8 Conclusion</b>	<b>21</b>
<b>Bibliography</b>	<b>23</b>
<b>A Appendix</b>	<b>25</b>



# 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>

---

Liv Arleth  
<????@student.aau.dk>

---

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



# 1. Introduction

## 1.1 Initial Problem Statement

### 1.1.1 Motivation

It is possible to use voice effects while performing. Many effects exist, and it is possible to change the parameters of an effect to one's liking in real time. A problem can be changing an effect and/or effect parameters while performing without having a big control board in front of the performer which can hinder 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?
  - What are the limits of the effects?
- Does any existing technology use body gestures or sensors to apply effects?

is rephrasing  
needed here?  
Olivier thought  
something should  
be done

### 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 arms or hands 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 arrises.

### 2.1 Research

This section focuses on the research necessary for the basic understanding of the subject.

[write some more](#)

#### 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, common effects will be explained.

##### 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, among a lot of other things.

##### 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 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 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. Firstly, the user chooses a reference of scales or tones, and secondly the amount correction to be made.

### **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 the 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.

finish this section  
with a short sum-  
mary/conclusion.

#### **2.1.2 State of the Art**

To understand and avoid issues a study of the state of the art on this area was conducted. There is a focus on commercial artifacts based on real-time alterations.

any comments on  
this intro?

#### **TC Helicon Perform V**

The TC Helicon - Perform V is a simple pedal 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 pedal effects are reverb, echo, "double" (harmonizer), EQ, compressor, and many more. It is possible to download an app that can connect with the Perform V. The application has many pre-made sounds, and it has a wireless connection.

## 2.1. Research



**Figure 2.1:** TC Helicon[7]

The Perform V is good for live performing if the singer has the pedal 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 singer limited to only three presets, and only one knob to turn.

### **Electro Harmonix Voice Box**

The Electro Harmonix Voice Box is a more advanced pedal 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. It is possible to insert an instrument to 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 mimu). 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, WiFi 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. It is possible to pre-order a pair of Mi.Mu gloves for £5,000, or one glove for £2,500. 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**

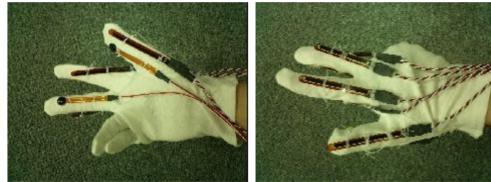
Yonezawa et al. made a glove that controls voice effects [10]. The wearer of the glove controls a puppet and makes hand gestures, see figure 2.4.



**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/opera, 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.

conclusion/summary  
here

#### **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 several ways to approach gesture control. For example, there are three stages of a gesture: registration, continuation, and termination, but in some cases the middle stage can be skipped [12, pp. 127-134]. Whether or not it is a good idea to skip this stage is decided by the number of different gestures one wants to implement. By removing the continuation stage, it removes the possibility to differentiate between a lot of gestures.

The registration stage is when the system registers that the user would want to gesture, 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

## 2.2. Problem Statement

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.

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

## 2.2 Problem Statement

## 2.3 Minimum Implementation

- The design must implement the use of an Arduino
- The design must implement the use of sensors applicable to the Arduino
- The design must implement audio processing
- The design must get audio from a microphone



### 3. Design



## 4. Implemented Theory



## 5. Implementation



## 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](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](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](http://www.gm.fh-koeln.de/hk/lehre/sgmci/ss2015/Literatur/Wigdor_Wixon_-_Brave_NUI_World.pdf), 2011.

## Bibliography

## A. Appendix