

---

---

# Speech Recognition

- With Information Retrieval & Command Prompt -

---

---

Project Report  
ED5-1-E17

Aalborg University  
Electronics and IT

Copyright © Aalborg University 2015

Here you can write something about which tools and software you have used for typesetting the document, running simulations and creating figures. If you do not know what to write, either leave this page blank or have a look at the colophon in some of your books.



**Electronics and IT**  
Aalborg University  
<http://www.aau.dk>

## **AALBORG UNIVERSITY**

### STUDENT REPORT

**Title:**

Speech Recognition

**Theme:**

Scientific Theme

**Project Period:**

Fall Semester 2017

**Project Group:**

ED5-1-E17

**Participant(s):**

Stefan Bîrs

Tiberiu-Ioan Szatmari

Kamran Thomas Alimagham

**Supervisor(s):**

Daniel Ortiz-Arroyo

D. M. Akbar Hussain

**Copies:** 1

**Page Numbers:** 11

**Date of Completion:**

September 19, 2017

**Abstract:**

This report goes into detail about the implementation of distributed speech recognition systems, information retrieval and machine learning. The main objective of the speech recognition system is to provide a handsfree human-computer interface experience, where the system is also able to handle noise.

*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>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Examples . . . . .	1
1.2 How Does Sections, Subsections, and Subsections Look? . . . . .	1
1.2.1 This is a Subsection . . . . .	1
<b>2 Chapter 2 name</b>	<b>3</b>
<b>3 Files</b>	<b>5</b>
3.1 .WAV . . . . .	5
<b>4 Machine Learning</b>	<b>7</b>
4.1 Speech Recognition with Deep Learning . . . . .	7
<b>5 Conclusion</b>	<b>9</b>
<b>A Appendix A name</b>	<b>11</b>



# Todo list

 Is it possible to add a subsubparagraph? . . . . .	2
 I think that a summary of this exciting chapter should be added. . . . .	2
 I think this word is misspelled . . . . .	3
Figure: We need a figure right here! . . . . .	3





# Preface

The project entitled Speech Recognition with Focus on information retrieval and command prompt was made by three students from the Electronics and Computer Engineering programme at Aalborg University Esbjerg, for the P5 project during the fifth semester.

Aalborg University, September 19, 2017

---

Stefan Birs

<sbars15@student.aau.dk>

---

Tiberiu-Ioan Szatmari

<tsz atm15@student.aau.dk>

---

Kamran Thomas Alimaghani

<kalima15@student.aau.dk>



# Chapter 1

## Introduction

Here is the introduction. The next chapter is chapter 2.  
a new paragraph

### 1.1 Examples

You can also have examples in your document such as in example 1.1.

#### **Example 1.1 (An Example of an Example)**

Here is an example with some math

$$0 = \exp(i\pi) + 1 . \tag{1.1}$$

You can adjust the colour and the line width in the `macros.tex` file.

### 1.2 How Does Sections, Subsections, and Subsections Look?

Well, like this

#### **1.2.1 This is a Subsection**

and this

#### **This is a Subsubsection**

and this.

**A Paragraph** You can also use paragraph titles which look like this.

**A Subparagraph** Moreover, you can also use subparagraph titles which look like this. They have a small indentation as opposed to the paragraph titles.

Is it possible to add a subsubparagraph?

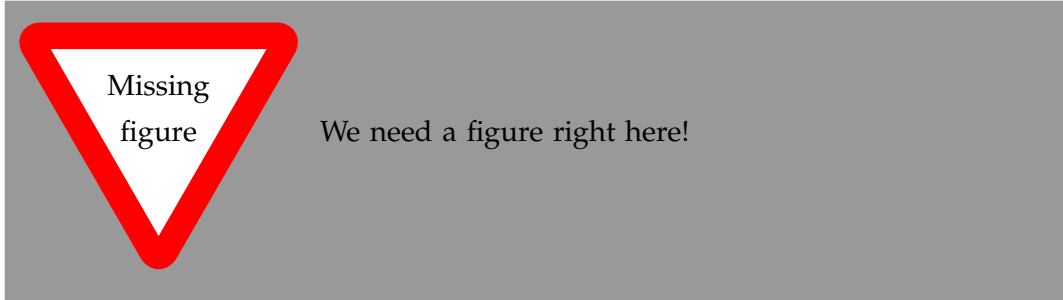
I think that a summary of this exciting chapter should be added.

## Chapter 2

## Chapter 2 name

Here is chapter 2. If you want to learn more about  $\text{\LaTeX}2_{\epsilon}$ , have a look at [Madsen2010], [Oetiker2010] and [Mittelbach2005].

I think this word is misspelled





## Chapter 3

# Files

### 3.1 .WAV

Waveform Audio File Format (WAVE, or more commonly known as WAV due to its filename extension)[3][6][7][8] (rarely, Audio for Windows)[9] is a Microsoft and IBM audio file format standard for storing an audio bitstream on PCs. It is an application of the Resource Interchange File Format (RIFF) bitstream format method for storing data in "chunks", and thus is also close to the 8SVX and the AIFF format used on Amiga and Macintosh computers, respectively. It is the main format used on Windows systems for raw and typically uncompressed audio. The usual bitstream encoding is the linear pulse-code modulation (LPCM) format.

Both WAVs and AIFFs are compatible with Windows, Macintosh, and Linux operating systems. The format takes into account some differences of the Intel CPU such as little-endian byte order. The RIFF format acts as a "wrapper" for various audio coding formats.

Though a WAV file can contain compressed audio, the most common WAV audio format is uncompressed audio in the linear pulse code modulation (LPCM) format. LPCM is also the standard audio coding format for audio CDs, which store two-channel LPCM audio sampled 44,100 times per second with 16 bits per sample. Since LPCM is uncompressed and retains all of the samples of an audio track, professional users or audio experts may use the WAV format with LPCM audio for maximum audio quality.[10] WAV files can also be edited and manipulated with relative ease using software.

The WAV format supports compressed audio, using, on Windows, the Audio Compression Manager. Any ACM codec can be used to compress a WAV file. The user interface (UI) for Audio Compression Manager may be accessed through various programs that use it, including Sound Recorder in some versions of Windows.

Beginning with Windows 2000, a WAVE\_FORMAT\_EXTENSIBLE header was de-

defined which specifies multiple audio channel data along with speaker positions, eliminates ambiguity regarding sample types and container sizes in the standard WAV format and supports defining custom extensions to the format chunk.[4][5][11] There are some inconsistencies in the WAV format: for example, 8-bit data is unsigned while 16-bit data is signed, and many chunks duplicate information found in other chunks.

WAV info - <<https://en.wikipedia.org/wiki/WAV>>

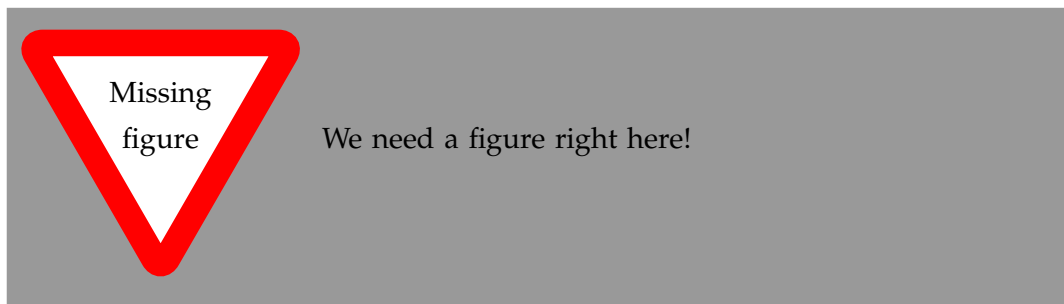


## Chapter 4

# Machine Learning

### 4.1 Speech Recognition with Deep Learning

Speech recognition is invading our lives. It's built into our phones, our game consoles and our smart watches. It's even automating our homes. For just \$50, you can get an Amazon Echo Dot a magic box that allows you to order pizza, get a weather report or even buy trash bags just by speaking out loud.



The Echo Dot has been so popular this holiday season that Amazon can't seem to keep them in stock! But speech recognition has been around for decades, so why is it just now hitting the mainstream? The reason is that deep learning finally made speech recognition accurate enough to be useful outside of carefully controlled environments. Andrew Ng has long predicted that as speech recognition goes from 95% accurate to 99% accurate, it will become a primary way that we interact with computers. The idea is that this 4% accuracy gap is the difference between annoyingly unreliable and incredibly useful. Thanks to Deep Learning, we're finally cresting that peak. Let's learn how to do speech recognition with deep learning!

speech recognition info - <https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to>



## Chapter 5

# Conclusion

In case you have questions, comments, suggestions or have found a bug, please do not hesitate to contact me. You can find my contact details below.

Jesper Kjær Nielsen  
jkn@es.aau.dk  
<http://kom.aau.dk/~jkn>  
Fredrik Bajers Vej 7  
9220 Aalborg Ø



## **Appendix A**

### **Appendix A name**

Here is the first appendix