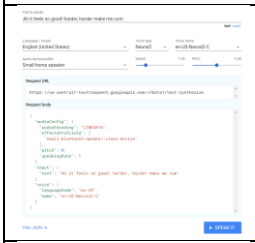
# Introduction

The aim of this week project is to develop a program that converts a pdf file to speech. Programs like this have been first created since 2015 with the Hunday car companies that created a system that read out loud messages to the vehicle driver using Google text-to-speech engine. (CNET. CBS Interactive. Retrieved 17 January 2015.)

Since this invention many other planforms have took on this idea such as companies like textPlus, WhatsApp, Google Cloud Text-to-Speech, Amazon and Microsoft. Google Cloud Text-to-Speech is powered by Wavenet a software created by Goodle UK based AI subsidirary DeepMind.

DeepMind AI voice synthesis tech is notably advanced and realistic. Most voice synthesizers (including Apple's Siri) use concatenative synthesis, in which a program stores individual phonemes and then pieces them together to form words and sentences. However, WaveNet speech that sounds more natural than other text-to-speech systems. It synthesizes speech with more human-like emphasis and inflection on syllables, phonemes, and words. Unlike most other text-to-speech systems, a WaveNet model creates raw audio waveforms from scratch. The model uses a neural network that has been trained using a large volume of speech samples. During training, the network extracts the underlying structure of the speech, such as which tones follow each other and what a realistic speech waveform looks like. When given a text input, the trained WaveNet model can generate the corresponding speech waveforms from scratch, one sample at a time, with up to 24,000 samples per second and seamless transitions between the individual sounds.

# Market Research

| Example | Name | Pros | Cons |
|---|---|---|---|
|  | Google Cloud Text-to-Speech | Very natural<br>Coherent<br>Suding voice<br>Ethnicities? | Lack of passion |
|  | Amazon ploy | Free to everyone,<br>Multiple tolls,<br>Can be used everywhere | Very robotic,<br>Souless |

| | | | |
|---|---|---|---|
|  | Microsoft Azaria | Have range of emotions,<br>Intonation,<br>Lots of options | Sounds word to word,<br>Pays 200 € |
|  | Speech -to-text (google default) | Can process lot of words | Very error prone |

## Target

Different examples of people interested in this game: **Who? What? Why** and **where?**

The type of persons interested in the software are:

Drivers that need to read text messages while driving, lazy people that don't like or are not able to read, people with troubled vison that impossibilities the reading, people with privacy issues that don't want to be recognise but want to share something or simple as part of a user machine interface to many other devices and products.

What? Must be able to read out loud text whiteout errors and in human structure to be to understandable.

Where? Places such as mobile devices, computers or tablets, on internet videos platform such as YouTube, on a portable device or leisure dulls.

## Design Specifications

The product requirement list (Table 2) gives me a clear information about important features that must be thought of when creating a product. It also showed me about the thought process required when developing the product.

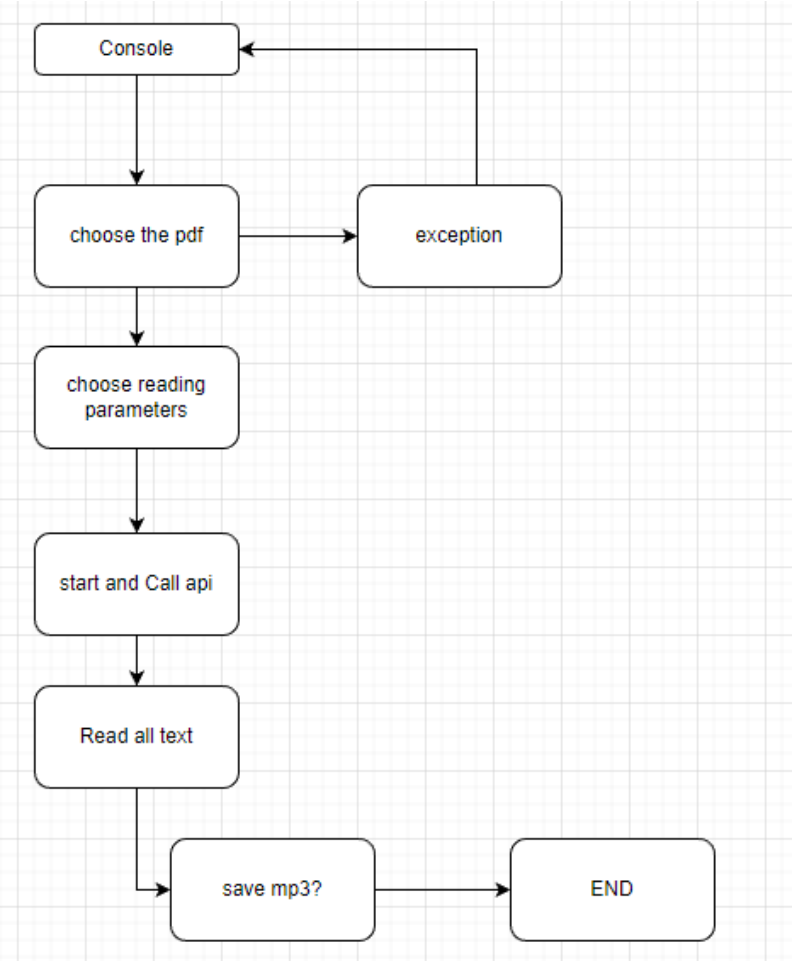
**D-** Stands for demands. These are requirements that are a must, they must be followed and executed when making sketches and coming up with ideas, such as health and safety requirements they are a must for all ideas to be developed.

**W**- Stands for wishes. These are requirements that, as a designer, we want our product to have.

| Requirements List<br>For the Breakout game | |
|---|---|
| D | *Can convert a text to speech* |

| | |
|---|---|
| W | *Easy to import text system* |
| D | *Uses a good Api for the reading* |
| D | *Reads pdf* |
| D | *Can read a whole book* |
| W | *Game over sequence with reset* |
| w | *Allows customisation* |
| w | *Reads other formats* |
| W | *Can safe to mp3* |
| W | *Can be made into device* |

## Flow Chart


Console
choose the pdf
exception
choose reading parameters
start and Call api
Read all text
save mp3?
END