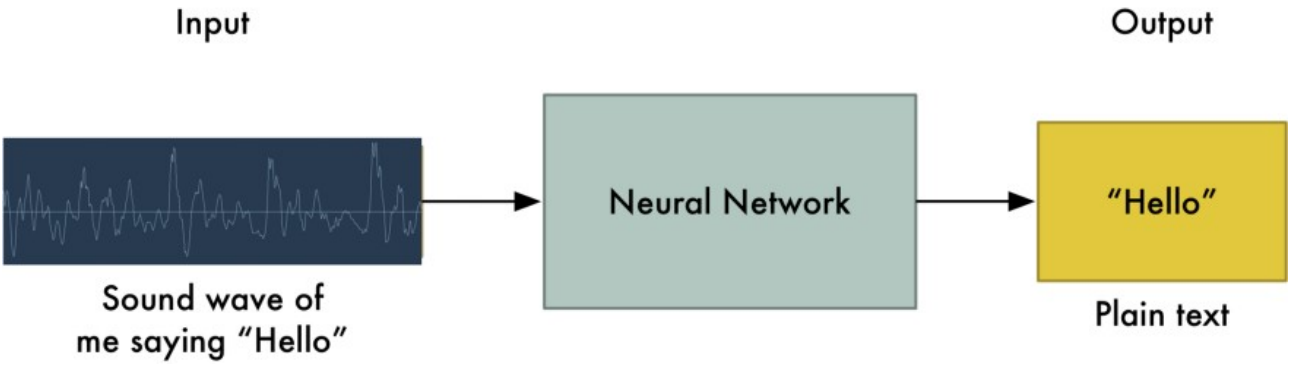
Speech Recognition

Introduction

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.

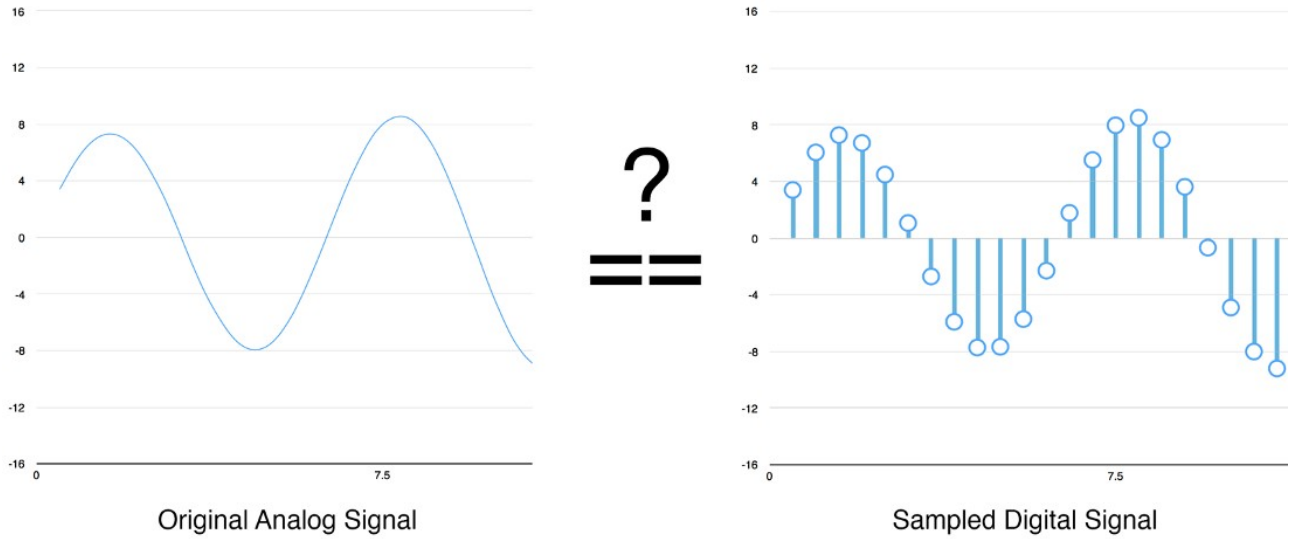
We could simply feed sound recordings into neural network and train it to produce text:



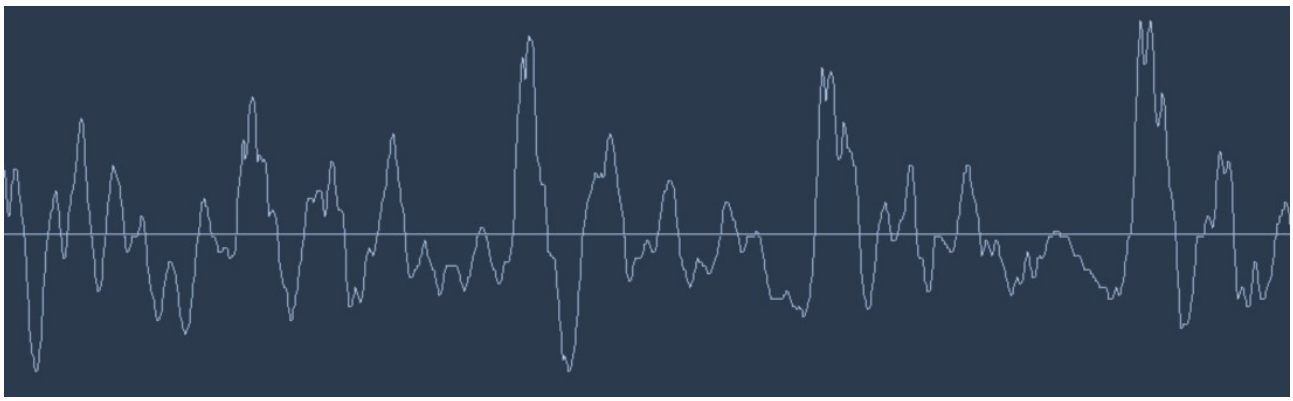
Although that is therotically all we have to do, the entire technique has certain limitations. People speak with different pace, with different speeds. Ones speech might sound like a ‘hello’ while others might sould like a ‘hellllllllo’. Despite of representing the same text hello, the two speech files might be significantly different.

Converting sound to bits

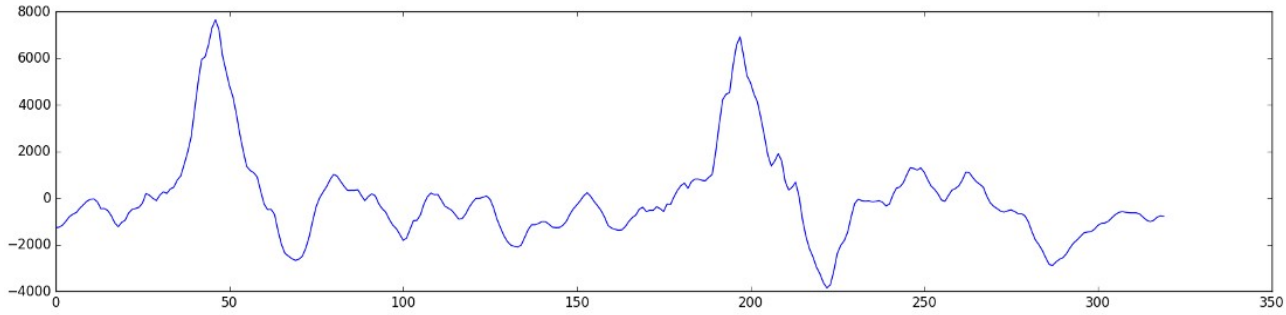
We can ‘sample’ the sound wave by recording the height of wave at equally spaced points. A 16 kHz audio should have been sampled 16,000 times and is enough to sample everything human could possibly hear. How reliable is the sampling technique? Surprisingly, 100%.



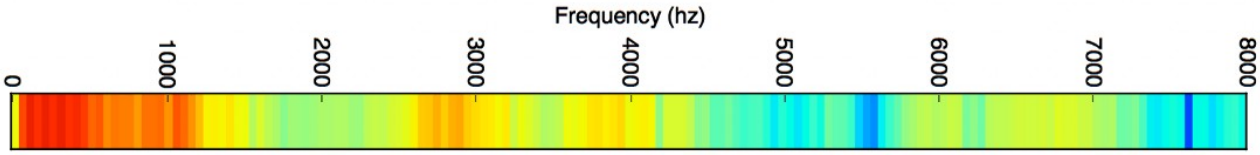
So, a waveform like this



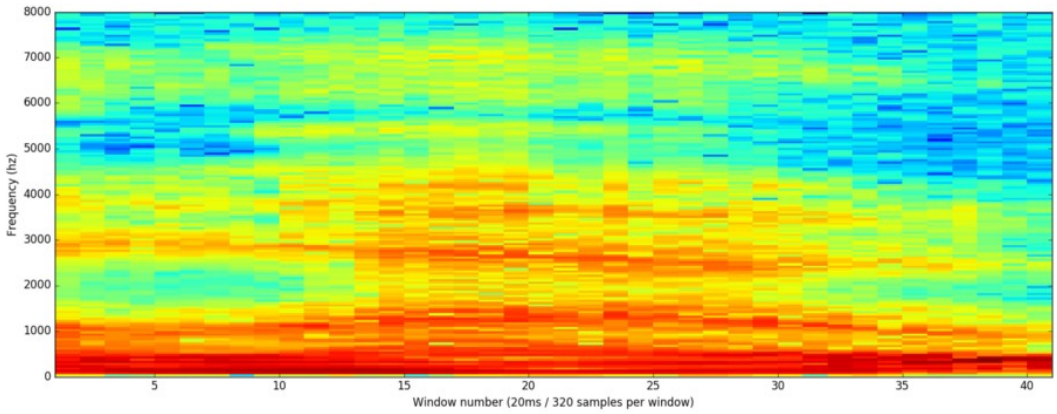
could be recorded as this. Please note that this is fraction of 20 ms of the entire clip.



Using some clever transformations, you can generate a score of how important each frequency range is, which looks like the following.



And when you repeat this every 20 ms, you get a spectogram.



Neural Networks, especially Recurrant Neural Networks are expected to

perform really well in classifying these patterns.

Our Proposal

Data generation and pre-processing: Generating dataset for numbers from 0 to 9 and pre-processing the data.

Training and testing: We would use the generated dataset for classifying numbers from 0 to 9.

Extension: We would try to incorporate natural way of saying numbers like 109 will not only be ‘one zero nine’ but also ‘one hundred nine’.

Natural Language Processing

Introduction

Natural Language processing is something which keeps artificial intelligence out of the race with Humans, but with artificial neural networks and powerful computing resources, Natural language processing is making life easier.

Our Proposal

Overall idea: A chatbot that would reply to questions. The answers won't be hard-coded like they used to be previously, but be learnt by the bot.

Data generation and pre-processing: It would be highly difficult to generate the huge amount of data that would be required to efficiently train a chatbot, so we would use already available data of dialogues of various movies and transform them to vectors.

Training and testing: The training part would be trivial. For testing, the way we are going to measure the accuracy has to be innovative. We haven't thought of it yet. One way is to manually test the chatbot, based on how reliable the response is and give a score manually. However this method is highly inefficient and unreliable.

Extension: This chatbot can be trained on customer service communication, this will make the chatbot a virtual assistant to those landing on your webpage.

Group members

111501029 Ashutosh Upadhye

111501023 Rahul Dhawan

111501028 Siddhardha SST