

IDC409 : Introduction to Data Sciences

# Speech Recognition System

GROUP - 6

Instructor : Dr. Vishal Bhardwaj

# Speech recognition systems

- Accurately translates spoken utterances into text
- Examples – Youtube closed captions, Siri front end , etc
- Importance of this system-
  1. Allow hands free control of devices
  2. Can act as a translator
  3. Aids the visually and hearing impaired
  4. Creates a print ready dictation
  5. Speeds up the process of documentation

# Why is it difficult to make this system?

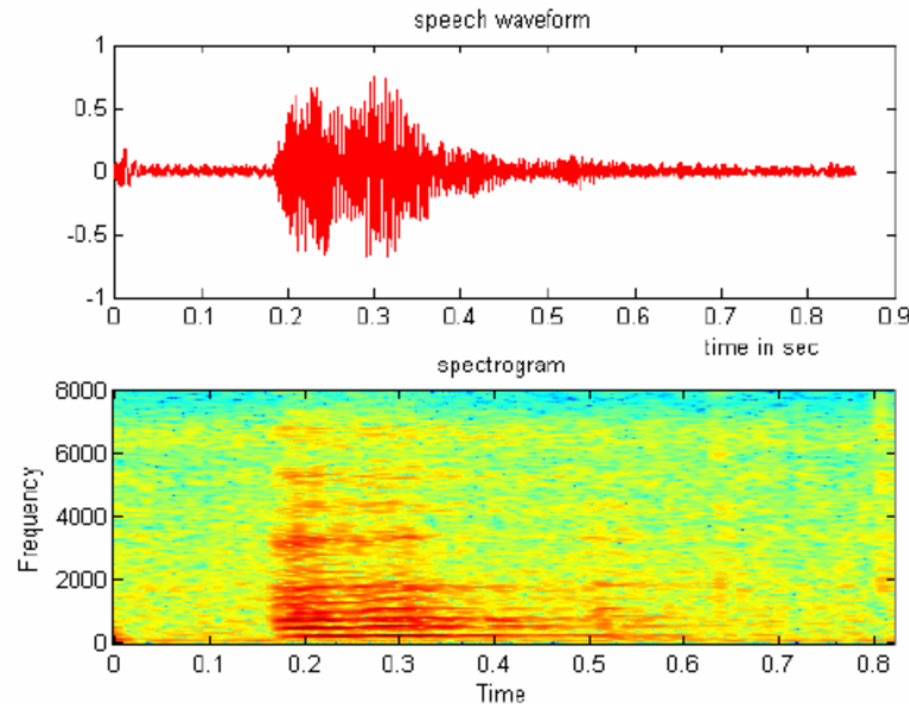
1. Style:
  - Continuous speech vs isolated words
2. Environment:
  - Background noise
3. Accent

# How does speech gets converted to text?



In the speech recognition process we will need three elements of sound. Those are →

1. Frequency
2. Intensity
3. The time it took to make it



We got the acoustic features!!!!

# Basic units of acoustic information

- Phoneme
  - a discrete unit or speech sound in a language which can be used to differentiate between words
  - 20-40ms long
  - Example- Four ----> f-ow-r

# Speech recognition systems

Speech recognition systems use two types of models:

- **Acoustic models.** These represent the relationship between linguistic units of speech and audio signals.

- **Language models.** Here, sounds are matched with word sequences to distinguish between words that sound similar.



## What are the different speech recognition algorithms?

The power behind speech recognition features comes from a set of algorithms and technologies. They include the following:

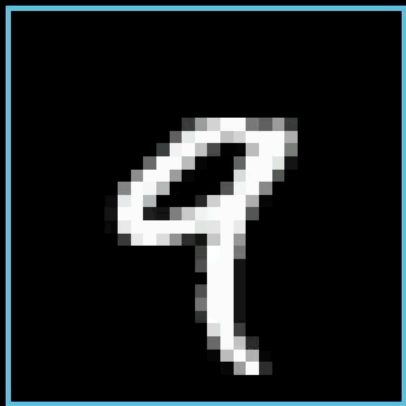
- Hidden Markov model.** [HMMs](#) are used in autonomous systems where a state is partially observable or when all of the information necessary to make a decision is not immediately available to the sensor (in speech recognition's case, a microphone). An example of this is in acoustic modeling, where a program must match linguistic units to audio signals using statistical probability.
- Natural language processing.** [NLP](#) eases and accelerates the speech recognition process.
- N-grams.** This simple approach to language models creates a probability distribution for a sequence. An example would be an algorithm that looks at the last few words spoken, approximates the history of the sample of speech and uses that to determine the probability of the next word or phrase that will be spoken.
- Artificial intelligence.** [AI](#) and [machine learning](#) methods like deep learning and neural networks are common in advanced speech recognition software. These systems use grammar, structure, syntax and composition of audio and voice signals to process speech. Machine learning systems gain knowledge with each use, making them well suited for nuances like accents.

Let's start with an example !

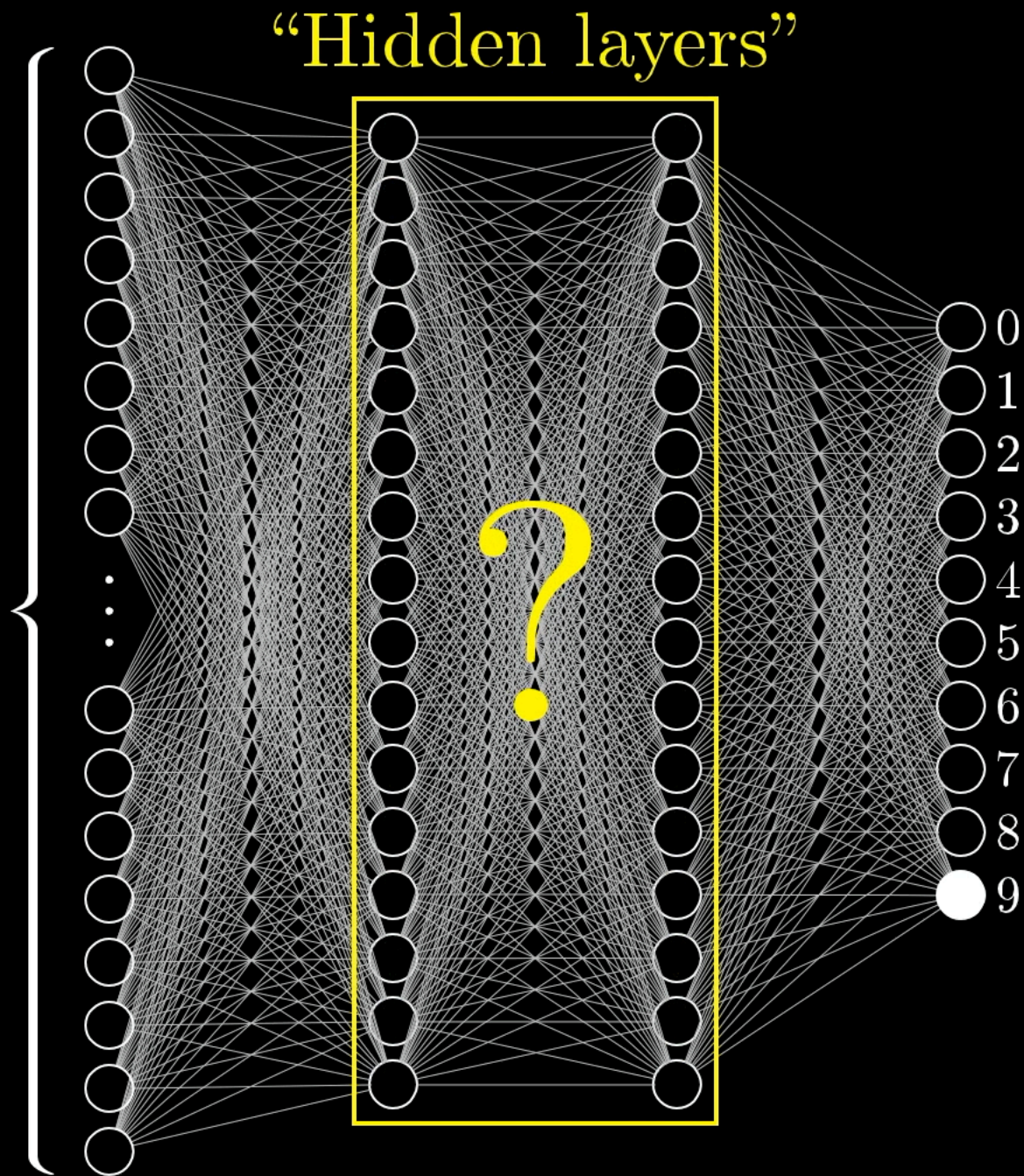


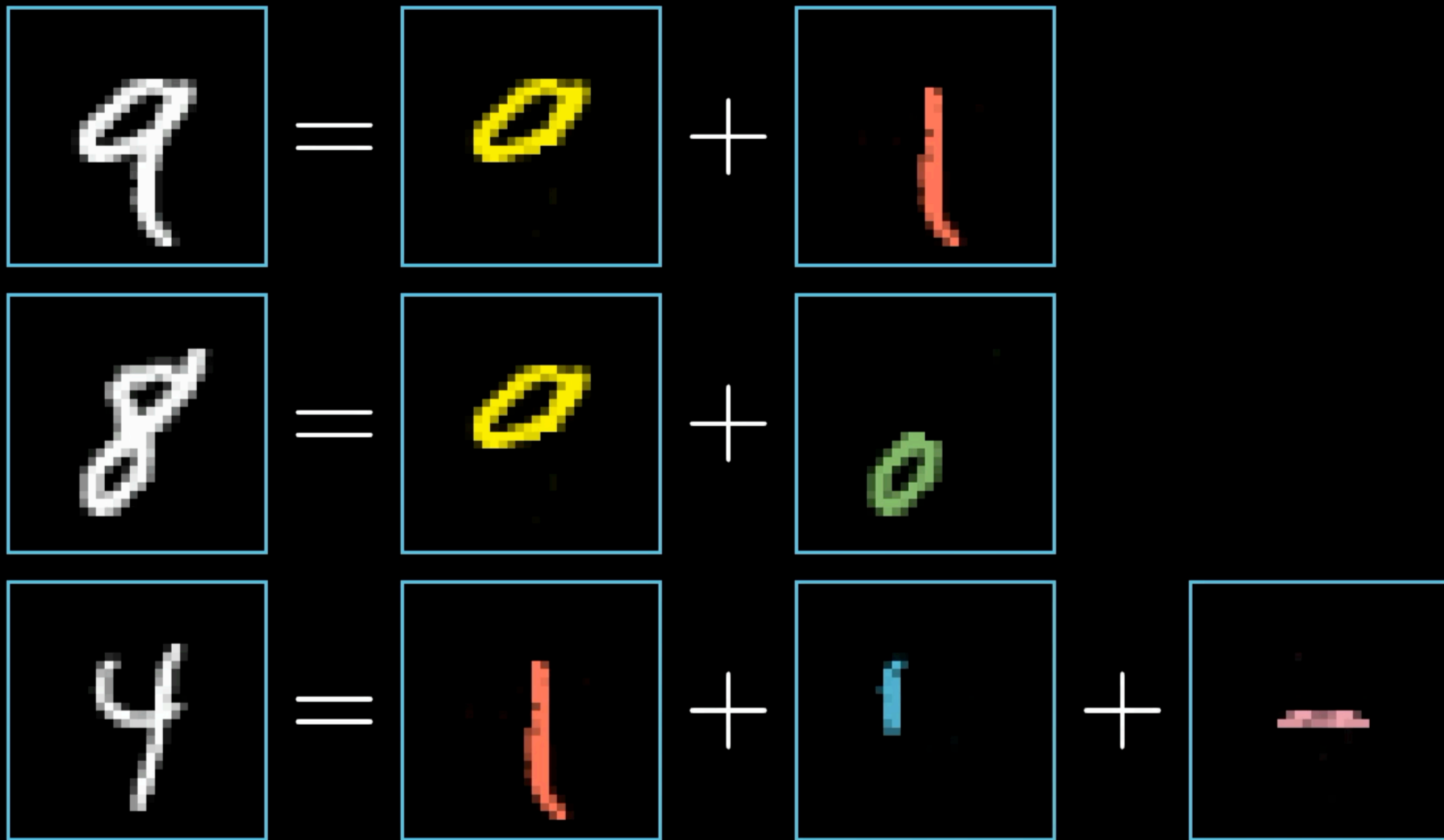
- There are many variants of neural networks, such as convolutional neural networks (CNN), recurrent neural networks (RNN), transformers, and countless others.
- In recent years there's been a boom in research of these variants. But the first step to understanding any of them is to build up the simplest, **plain vanilla** form with no added frills.-aka **multilayer perception**

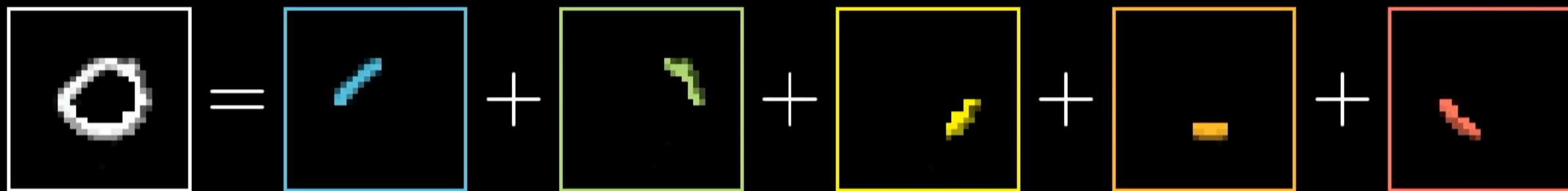




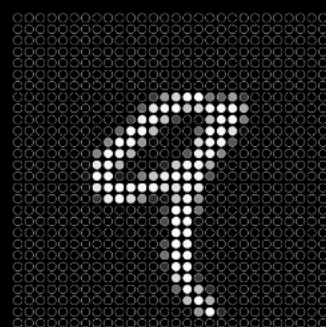
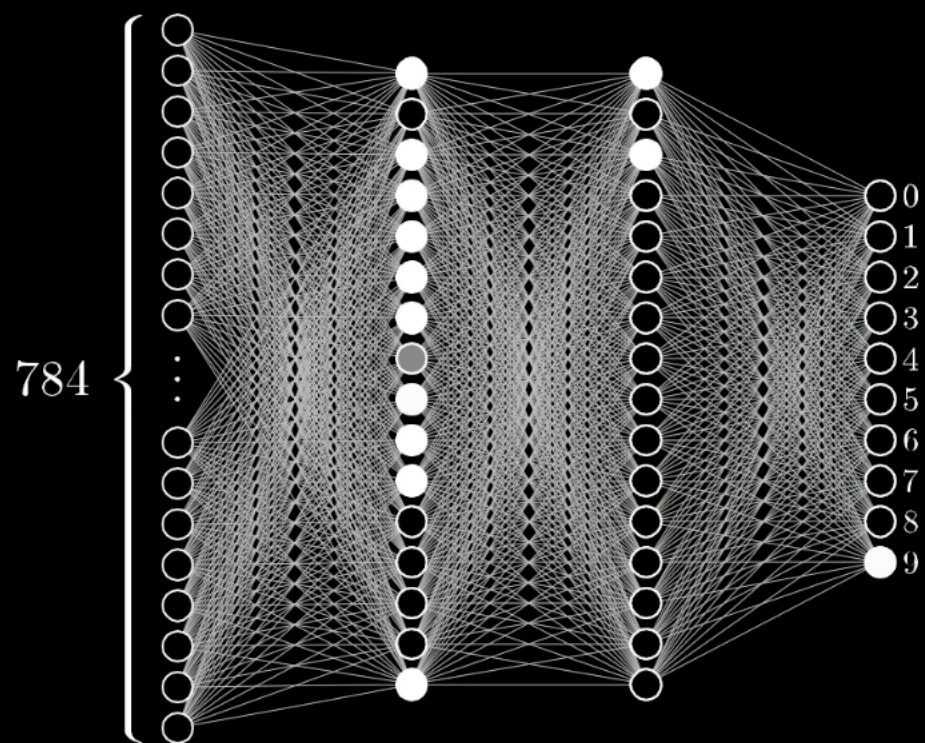
784











Pixels



Edges

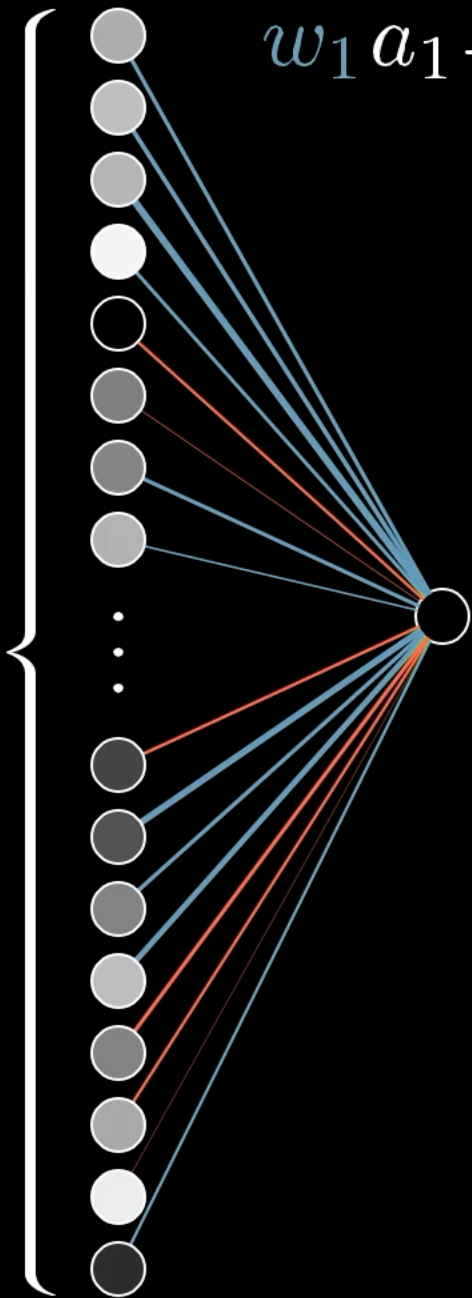


Shapes

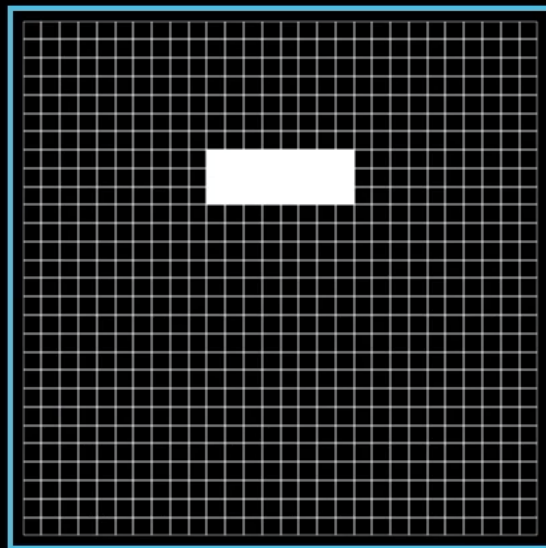
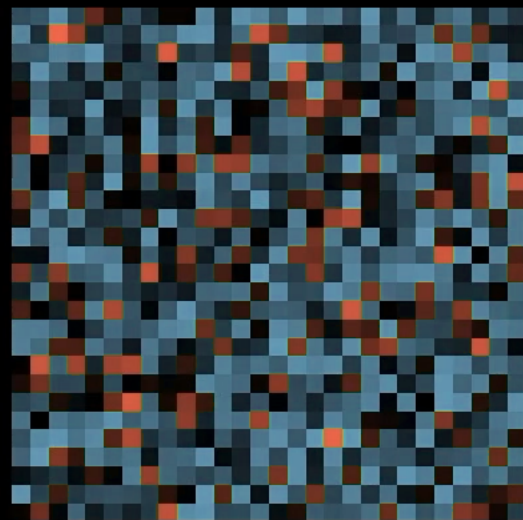


Digits

784



$$w_1 a_1 + w_2 a_2 + w_3 a_3 + w_4 a_4 + \cdots + w_n a_n$$





Raw audio



Raw audio



Raw audio



r e c o g n i t i o n



re·cog·ni·tion



recognition

Raw audio

# Future Prospects:

The prospects of speech recognition systems are quite promising. Here are some key features and developments to expect:

1. Improved Accuracy
2. Multilingual Support
3. Context Awareness
4. Integration with IoT
5. Personalization
6. Accessibility
7. Real-time Translation
8. Industry-Specific Applications
9. Security
10. Natural Interaction
11. Offline Functionality
12. Noise Cancellation
13. Emotional Analysis

## Some References

- <https://youtu.be/ZqpSb5p1xQo?si=weR7LF2jAH5xl8Zs>
- [https://youtu.be/U0XtE4\\_QLXI?si=0gw1q\\_rO2WUXj3Va](https://youtu.be/U0XtE4_QLXI?si=0gw1q_rO2WUXj3Va)
- <https://youtu.be/aircAruvnKk?si=63yuuxHt19lexNnb>
- <https://www.3blue1brown.com/topics/neural-networks>
- <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>
- <https://distill.pub/2017/ctc/>



**THANK YOU !!**