**Artificial Intelligence Speech Recognition and History**

Joshua J. Gearhart

Slippery Rock University – Student in Computer Science and Cybersecurity

CLSD – 244 - Artificial intelligence and Films

Dr. Michael J. Ardoline

May 6th, 2022.

1. **Abstract**

This paper will review the advancements of artificial intelligence within speech recognition, as well as its history. This paper will review methods that are being implemented to improve Artificial Intelligence and their capabilities for handling speech. This will also include how they process it, understand it, as well as how they can respond to it. The goal of this paper is to create a cumulative knowledge of artificial intelligence and learn about how their speech could operate.

1. **Introduction**

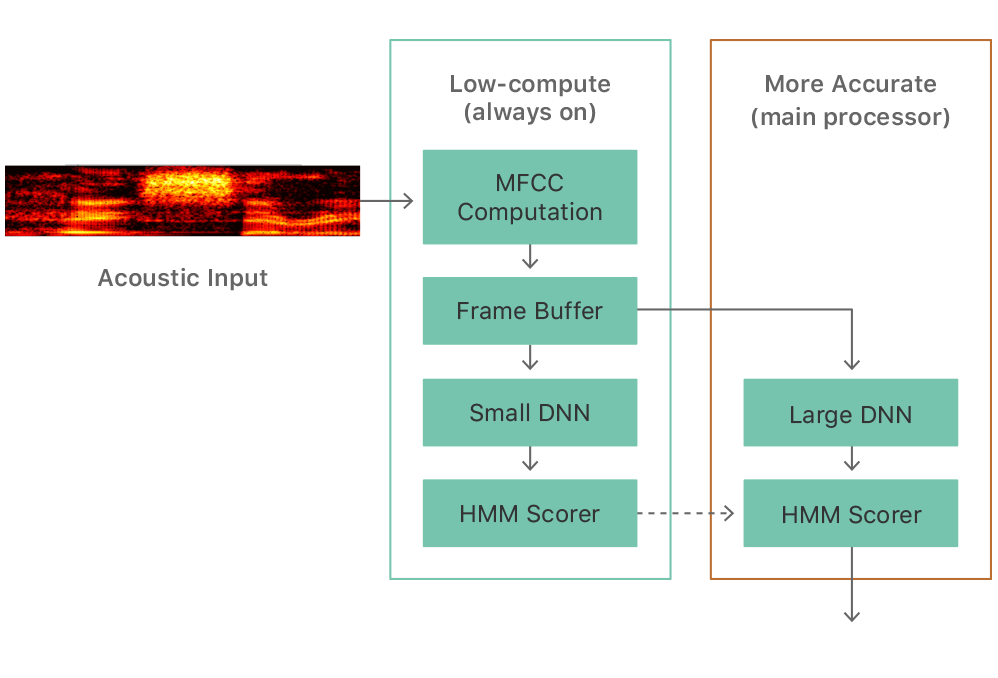
Artificial Intelligence, or what will be referred to as A.I. in this paper has always been an exciting and “futuristic” topic in our world; the idea of machines and having them learn things is quite fascinating to us as intelligent beings. Our idea of teaching a machine to recognize speech has quite a deep history, starting in the 1950s, the first A.I. known to recognize human speech is attributed to the creators at Bell Laboratories. They created “Audrey” or her functional name “automatic digit recognizer” in 1952.[1] Her job was to receive digits from a human speaker with spaces in between each number, she could understand zero to nine with a reported ninety percent accuracy which is quite impressive for the time. Though this was the first time A.I. had been able to understand humans, it was quite early into the field of computer science and Audrey occupied a six-foot-tall server rack[5], this was in the 50s so a computer required to run her hardware could probably be found to run in any smartphone today. In this paper, we will look at the advancements of A.I. Speech recognition from the least known Audrey all the way to our iPhone Siri and what our future could look like.

1. **Methods**

The methods used to write this paper were many literature reviews, reviewing, and trying to learn about a deep and complex structure of acoustic analysis to understand how much actually goes into speech recognition. This paper is a combination of the history of speech recognition and what is currently being used today.

1. **Discussion**

After Audrey’s creation came the next voice recognition system where we saw a top computing company today create the next step in history, this was done by IBM, who created “Shoebox” in the 60s. Shoebox was basically a more advanced Audrey, they were similar in that they could understand digits zero to nine however they differed in the fact that Shoebox could also understand sixteen spoken words. These words included things like plus, minus, and total. This allowed Shoebox to communicate with other machines to print and run simple mathematics. [3] In the 70s came the Harpy system created by Carnegie Mellon’s “Harpy” Speech system this was funded by the United States department of defense. [1] Once it came to the 80s an even bigger step to A.I. Speech recognition came to light, this was when a statistical algorithm was used on speech recognition, specifically, this algorithm works with patterns. This is called the Hidden Markov Model, it is used to recognize words, it’s important to note that this algorithm is “trained” which means it has inputs to base off. Another algorithm is used in combination with the HMM model. This algorithm creates frames of each “utterance” or word to analyze acoustically, the word to be recognized than is completely searched on a vector matrix with a search algorithm and is compared, it’s given a score comparatively to the trained words, and the word is then matched with the shortest pathed word. [4] This model became the standard for speech recognition. As speech recognition has advanced the use of HMM and how it operates has as well, this is attributed to not only Moore’s Law, but also to the increase in algorithmic models in combination with speech recognition. Currently, big companies use neural networks powered by data centers and trained by searches that the software uses. I am going to talk about Siri and how she “wakes up” by the phrase “Hey Siri”. Siri is always listening when this is enabled, however, she only responds when in a sense spoken to. This is done by using what’s known as a neural network which is a shared data model, where each node communicates with the other and so on. How this works is that when a secondary lower-powered processor with direct access to the iPhone microphone detects words, it’s scored with the HMM model, if it scores above a threshold value then the bigger processor wakes up to receive and process it to handle the speech. Below is Fig. 1. depicting this process within an iPhone. [5]



**Figure 1.**

How Siri detects the “Hey Siri” and then works to take in audio input.[5]

1. **Conclusion**

We have discussed the history of speech recognition, its advancements of it, and how we are where we are today. I believe in the future we will see better improvements, though it may seem we have it all figured out we aren’t perfect yet, models and algorithms can always be tweaked to improve efficiency, and I believe neural networks and advancements within that technology will also grant us further improvements.

1. **References**
2. *A brief history of speech recognition*. Sonix. (n.d.). Retrieved April 30, 2022, from <https://sonix.ai/history-of-speech-recognition>
3. *Hey Siri: An on-device DNN-powered voice trigger for Apple's personal assistant*. Apple Machine Learning Research. (n.d.). Retrieved May 1, 2022, from <https://machinelearning.apple.com/research/hey-siri>
4. *IBM shoebox*. IBM Archives: IBM Shoebox. (n.d.). Retrieved May 5, 2022, from <https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html>
5. Najkar, N., Razzazi, F., & Sameti, H. (2010, March 25). *A novel approach to HMM-based speech recognition systems using particle swarm optimization*. Mathematical and Computer Modelling. Retrieved May 5, 2022, from <https://www.sciencedirect.com/science/article/pii/S0895717710001597#fig1>
6. Warren, J. (2014, October 13). *Audrey: The first speech recognition system*. Asta Speaks. Retrieved May 3, 2022, from <https://astaspeaks.wordpress.com/2014/10/13/audrey-the-first-speech-recognition-system/>