LOW POWER KEYWORD SPOTTING

JOHN-ANTHONY ELENIS • DALJIT JOSH • HEMAN MURESAN

WAVE TECH

Problem Statement

IoT devices actively run expensive speech recognition algorithms when no audio may be present, this reduces the total battery life and passes costs directly onto the consumer.

Background

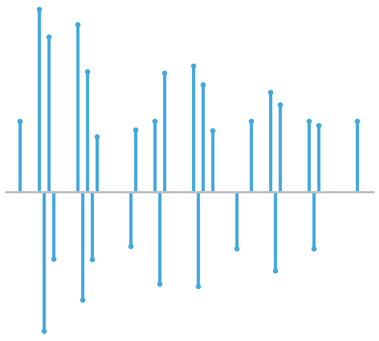
Recent improvements in cloud computing and the computational power of embedded devices have fostered a growing industry of smart home technology. Smart speakers continually listen for a phrase before they begin processing audio. This technique is commonly referred to as keyword spotting.

Here, we take a multistage approach to keyword spotting. This means that the more power intensive operations are avoided until completion of the prior stage. This saves power as the complexity grows in each successive stage.

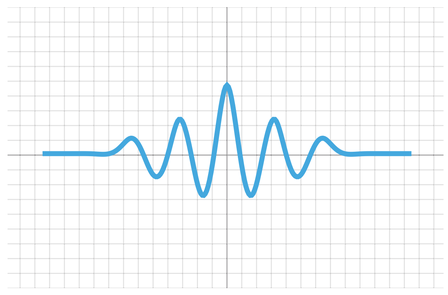
Design Goals

- Introduce a multistage system that isolates power consuming tasks
- Small form factor to easily integrate with existing IoT solutions
- Increase power efficiency without loss of precision

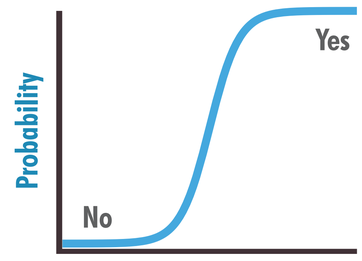
Speech



Sampled Audio



Voice Detection

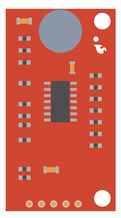


Keyword Detection

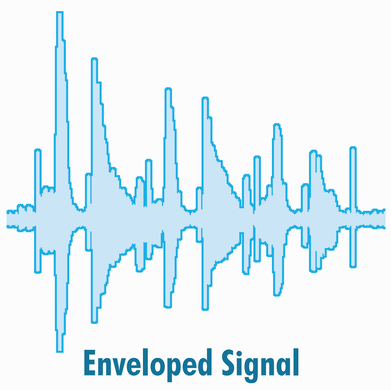
Power Consumption

Audio Sampling

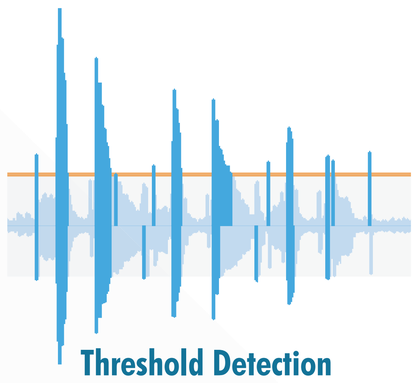
The first stage is responsible for filtering out ambient noise and creating an approximate digital signal from the analog audio. This is performed by creating an envelope around the amplified analog signal and only taking samples when the audio envelope is greater than a threshold.



Low Power MIC



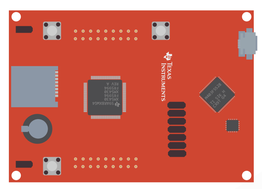
Enveloped Signal



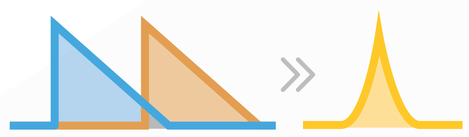
Threshold Detection

Voice Detection

This stage is responsible for verifying if human voice is present in the audio signal. Autocorrelation is used to observe the fundamental frequency of the signal which is then compared with the known frequency range of human voices.



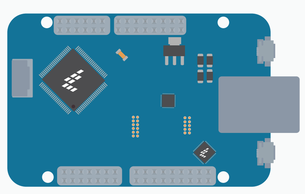
Low Powered MCU



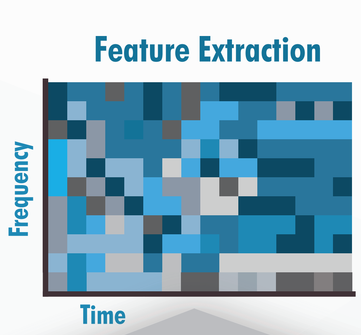
Autocorrelation

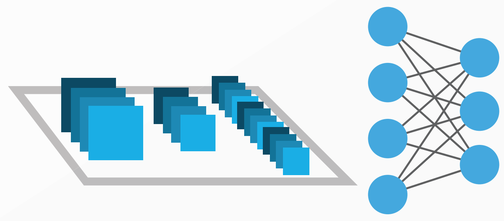
Keyword Detection

Keywords are detected by extracting the frequencies from the spoken audio and using them as input features to the neural network. Since the network is tuned specifically for this pattern of frequencies, it can give a probability of a keyword's likeness.



High Power MCU





Convolutional Neural Network

Conclusion

By reducing the frequency of power intensive operations, this multistage approach provides a means to significantly reduce total power consumption. It was found that power savings can be further improved by raising the threshold during audio sampling and reducing the systems sensitivity to voices. However, this is a trade off as it will come at the expense of negatively affecting performance.

UNIVERSITY of GUELPH
ENGINEERING
Faculty Advisor
STEFANO GREGORI, PhD. P.Eng
Computer Systems and Computing