

ENGI 301

VocalVision Proposal


Speaker Identification and Transcription

Sept. 26, 2023

PRESENTED BY
Parnika Mandewalkar

Background Information

What is Vocal Vision?



VocalVision is a **real-time speech interpreter** that can listen into meetings or hard-to-follow conversations using one mic and **recognize who is speaking as well as transcribe** what they are saying onto a display.

Implications

This device can have a majority of applications, the intended one being aiding deaf people in the interpretation of large-scale in-person meetings. Generally, it is immensely difficult to follow who is speaking and what they are saying when there are so many people participating in a conversation, so visualizing this onto a display can be of immense help.

What would this involve?

Lots of machine learning and data analysis. A mic would be connected to the PocketBeagle to obtain real time data and it would output onto an LCD display. The python libraries used would be tensorflow, keras, librosa, etc. Vocal Classification exists, and so does transcription, but not together just yet!

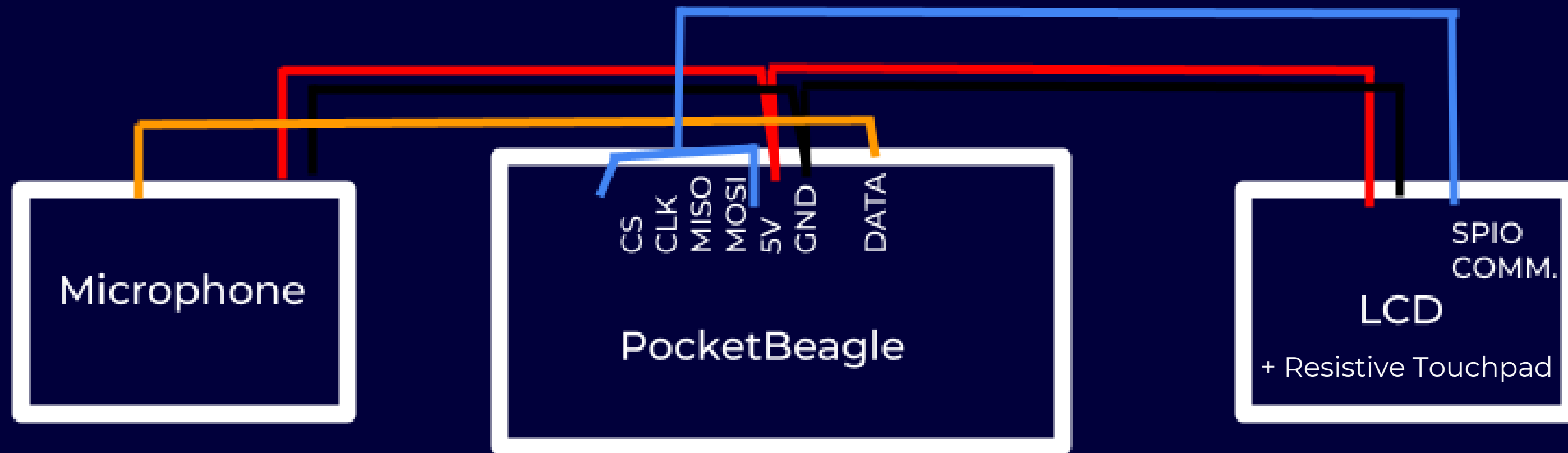
References

<https://towardsdatascience.com/voice-classification-using-deep-learning-with-python-6eddb9580381>

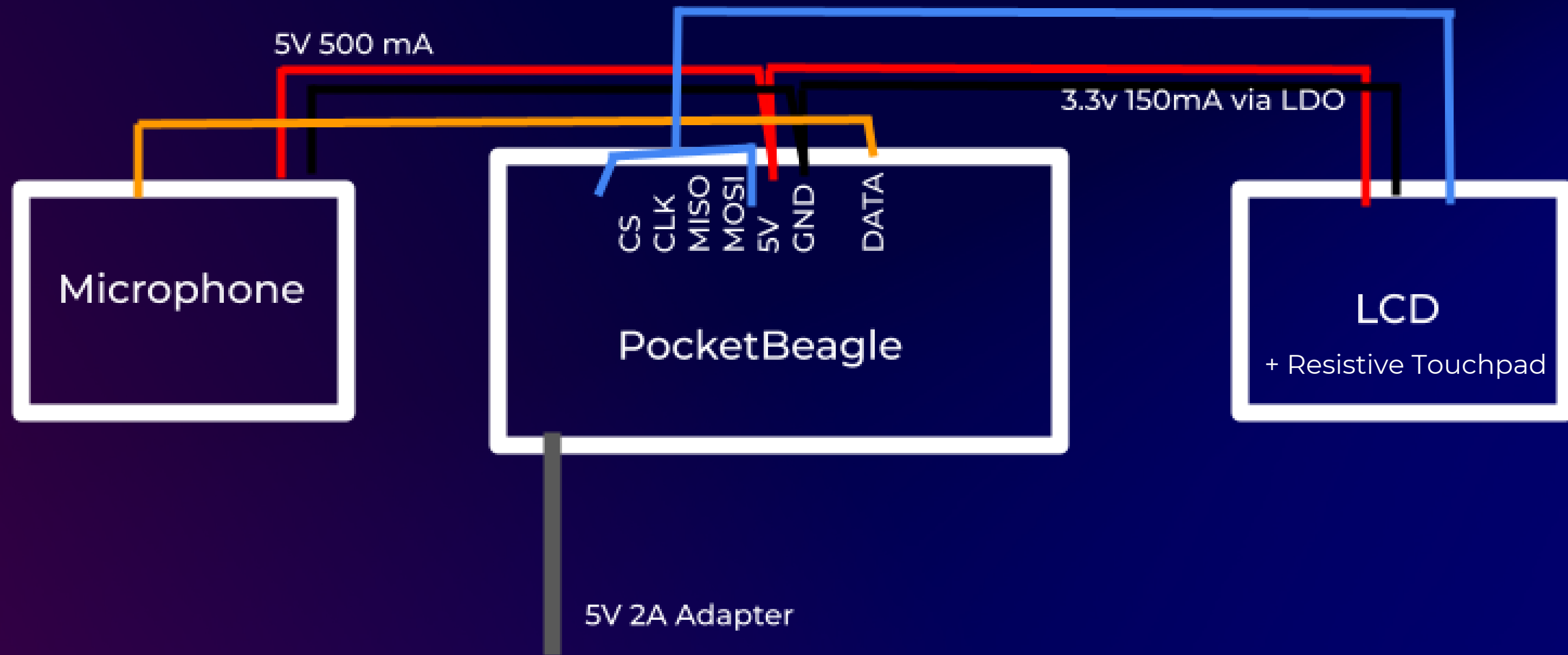
<https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5>

PocketBeagle Alexa Project (For input audio analysis)

System Block Diagram



Power Block Diagram



Components/Budget

Part		Value	Link	Buy?
PocketBeagle		-	-	
2.8" TFT LCD - ILI9341 (Parts List)		\$29.95	https://www.adafruit.com/product/1770	✓
USB 2.0 Microphone (Parts List)		\$7.99	https://www.amazon.com/gp/product/B01MQ2AA0X	✓
Resistive Touch Controller (Parts List)		\$4.95	https://www.adafruit.com/product/5423	✓
TOTAL		\$42.89		